WORKING PAPERS

RESEARCH DEPARTMENT

WORKING PAPER NO. 16-11
A NARRATIVE APPROACH TO A FISCAL DSGE MODEL

Thorsten Drautzburg
Research Department
Federal Reserve Bank of Philadelphia

April 2016

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

A Narrative Approach to a Fiscal DSGE Model

This version: March 28, 2016

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 based on narrative shocks. This paper asks whether both approaches agree. First, I show that, theoretically, the narrative VAR approach is valid in a class of DSGE models with Taylor-type policy rules. Second, I quantify whether the two approaches also agree empirically, that is, whether DSGE model restrictions on the VARs and the narrative variables are supported by the data. To that end, I first adapt the existing methods for shock identification with external instruments for Bayesian VARs in the SUR framework. I also extend the DSGE-VAR framework to incorporate these instruments. Based on a standard DSGE model with fiscal rules, my results indicate that the DSGE model identification is at odds with the narrative information as measured by the marginal likelihood. I trace this discrepancy to differences both in impulse responses and identified historical shocks.

Keywords: fiscal policy, monetary policy, DSGE model, Bayesian estimation, narrative shocks, Bayesian VAR

JEL Classifications: C32, E32, E52, E62

* Federal Reserve Bank of Philadelphia. E-mail: tdrautzburg[at]gmail.com. Web: sites.google.com/site/tdrautzburg/. I am grateful to Pooyan Amir-Ahmadi, Jesus Fernandez-Villaverde, Pablo Guerron-Quintana, my discussant Ed Herbst, Jim Nason, Barbara Rossi, Juan Rubio-Ramirez, Frank Schorfheide, Tom Stark, Keith Sill, Harald Uhlig, Mark Watson, Jonathan Wright, and audiences at the Atlanta Fed, the 2014 EEA, the ECB, the 2015 ESWC, Emory, Goethe, Madison, the 2015 NBER-NSF SBIES and NBER SI, the Philadelphia Fed, the Riksbank, the 2014 SED, the Fall 2015 System Macro Committee, and Wharton for comments and suggestions. I would also like to thank Valerie Ramey, Karel Mertens, and Morten Ravn for making their code and data publicly available. All errors are mine.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are a widespread research and policy tool. But as structural models, DSGE models could be misspecified and their results therefore misleading. An alternative approach uses narrative methods to study some of the same shocks and their effects. Narrative methods rely on fewer structural assumptions and incorporate additional information relative to standard macroeconomic time series. I provide a framework for incorporating this information in DSGE model estimation and to quantify misspecification compared with narrative studies.

Despite the success of DSGE models such as, for example, Christiano et al. (2005) and Smets and Wouters (2007), concerns about misspecification remain: Faust (2009), for example, argues that microfoundations of DSGE models are weak and that their quantitative success could mask misspecification. Sims (2005) cautions that DSGE models should not displace alternative identification schemes. Del Negro et al. (2007) use the Del Negro and Schorfheide (2004) framework to quantify model misspecification relative to a reduced-form vector autoregression (VAR) and find a small degree of misspecification. But since their benchmark model is a reduced-form VAR, their metric for comparing models is unaffected by shock identification.[1]

In this paper, I quantify the misspecification of DSGE models with a focus on shock identification. I extend the Del Negro and Schorfheide (2004) framework for assessing misspecification in reduced-form VARs to structural VARs identified with external instruments: I model traditional macro variables jointly with narrative shocks such as the monetary policy shocks developed by Romer and Romer (2004). With the resulting narrative DSGE-VAR, I quantify the misspecification of the DSGE identification scheme by comparing marginal likelihoods. If a given DSGE model and the narrative VAR agree, including information from the model increases precision and hence the marginal likelihood.[2,3]

Comparing DSGE models with narrative VARs is valid for a class of DSGE models: I provide conditions under which the instrument-identified VAR in Mertens and Ravn (2013) correctly identifies shocks and policy rule coefficients in models with standard Taylor-type policy rules, such as Leeper et al. (2010) and Fernandez-Villaverde et al. (2015), without timing restrictions. This property of the narrative VAR contrasts with traditional VARs that identify shocks through contemporaneous zero restrictions. DSGE models that match the VAR then need to assume that economic agents only react to policy shocks with a delay. The narrative VAR approach is valid if the data are generated from a widely used class of DSGE models, without restricting the timing.

My application focuses on fiscal and monetary policy rules in a medium-scale DSGE model, with narrative measures for government spending, tax rates, and monetary policy shocks. I find that the DSGE model augmented with Taylor-type policy rules agrees with narrative VARs as measured by marginal likelihoods: The marginal likelihood is initially increasing in the prior weight on the DSGE

[1] Canova et al. (2015) use the Del Negro and Schorfheide (2004) DSGE-VAR method to assess whether the assumption of time-invariant parameters matters.

[2] Formally, shocks in the standard DSGE-VAR framework are identified only up to multiplication by an orthonormal matrix that does not affect the likelihood. Here, the extra data, and hence the likelihood, also speak to the identification.

[3] Waggoner and Zha (2012) confront model misspecification differently. Instead of estimating a constant mixture between the DSGE model and the VAR, they allow for Markov switching between the models. Given the limited numbers of narrative shock observations, such a Markov switching approach seems too demanding of the data in my application.

model. While the marginal likelihood peaks at an interior point in terms of the DSGE model weight, it is decreasing in the weight put on the DSGE covariance matrix and, hence, the identification scheme. In contrast, shrinkage toward the DSGE model dynamics improves the model fit initially. I interpret these results as indicating that the DSGE model identification does not line up with the narrative identification. These results are robust to estimating the model separately with the monetary or fiscal instruments. To provide evidence on the specific channels, I also compare impulse response functions (IRFs) and historical shocks. Regarding the historical shocks, the DSGE model does best at matching up the monetary policy shocks that the narrative VAR implies and worst for the tax shocks. The model also struggles particularly to explain the inflation and tax dynamics in the narrative VAR.

Since it provides an intuitive way to incorporate prior information, the narrative DSGE-VAR framework can also be of interest for future narrative studies. In my application, the posterior distributions over the effects of fiscal shocks remain wide based on narrative shocks alone. Incorporating prior information sharpens the posterior by shrinking the estimates toward theory-consistent policy rules. This is similar to the work by Arias et al. (2015), who document that even weak structural priors over policy rules help to identify shocks via sign restrictions.

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. (2011b) 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 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.[4]

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. A web appendix contains the proofs and additional 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:

$$Y_t = B^* X^*_{t-1} + A^* \epsilon^*_t \tag{2.1a}$$

[4]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) imply 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.

$$X_t^* = D^* X_{t-1}^* + C^* \epsilon_t^*, \tag{2.1b}$$

where $\mathbb{E}[\epsilon_t^*(\epsilon_t^*)'] = I_m$. 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:

$$Y_t = BX_{t-1} + A\epsilon_t \tag{2.2a}$$

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

where the dimension of the state vector is typically different across the VAR and the DSGE model but the shocks ϵ_t and ϵ^*_t are assumed to be of the same dimension. Importantly, the observables Y_t are also the same.

The typical challenge in VARs is that observables only identify AA' but not A. The standard DSGE-VAR in Del Negro et al. (2007) therefore uses the same rotation scheme that identifies shocks in the DSGE model also to identify shocks in the VAR. In contrast, I bring in additional data to identify a subset of shocks in the VAR.

Stock and Watson (2012) and Mertens and Ravn (2013) have shown how to use narrative or external instruments to (partially) identify A for given parameter estimates. In the next section, I first summarize their identification results and show how to use a standard seemingly unrelated regression (SUR) estimator for inference. Second, I provide conditions when this procedure is correct in a class of DSGE models. Last, I show how to implement a prior for the narrative VAR based on a DSGE model.

3 Bayesian VAR estimation with narrative instruments

In this section, I first discuss identification and Bayesian estimation in the instrument-identified VAR. I refer to this model as narrative BVAR in what follows. 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.[5]

3.1 Narrative BVAR

I use the following notation for the statistical model:

$$y_t = \mu_y + By_{t-1} + v_t \tag{3.1a}$$

$$v_t = A\epsilon_t^{str}, \epsilon_t^{str} \overset{iid}{\sim} \mathcal{N}(0, I_m) \tag{3.1b}$$

$$z_t = \mu_z + Fv_t + \Omega^{-1/2}u_t, u_t \overset{iid}{\sim} \mathcal{N}(0, I_k) \tag{3.1c}$$

[5] Caldara and Herbst (2015) provide an alternative way to estimate a narrative or proxy VAR using Bayesian methods. Their approach factors the likelihood function differently than I do and thereby highlights nicely how the instruments restrict the rotation matrix that maps forecast errors to structural shocks.

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.[6]

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:

$$z_t = \begin{bmatrix} G & 0 \end{bmatrix} \epsilon_t^{str} + \Omega^{-1/2} u_t = \underbrace{\begin{bmatrix} G & 0 \end{bmatrix} A^{-1}}_{\equiv F} \underbrace{A\epsilon_t^{str}}_{\equiv v_t} + \Omega^{-1/2} u_t \tag{3.2}$$

By imposing zero restrictions on the structural representation of the covariance matrix, knowledge of F and AA' identifies the shocks that are not included in (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 Γ can be written as:*

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

The assumption that G is invertible follows Mertens and Ravn (2013) and corresponds to the assumption that the instruments are relevant.

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

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

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

Using the definitions of Γ and the forecast errors gives from (3.3):

$$\Gamma \equiv \text{Cov}[z_t, v_t] = \text{Cov}[z_t, A\epsilon_t^{str}] = \begin{bmatrix} G & 0 \end{bmatrix} A' = G\alpha_1' = [G\alpha_{11}', G\alpha_{21}'] \tag{3.5}$$

[6] One could easily generalize the model to include conditioning information in the instrument equation. For example, I use forward-looking information from Ramey (2011) and Fisher and Peters (2010) in robustness checks to make sure the instrument does not pick up the measure anticipatory policy measures.

Under Assumption 1, G is invertible, and we can partition Γ so that $G^{-1} = \alpha'_{11}\Gamma_1^{-1}$. From (3.5), we also know that $\alpha'_{21} = G^{-1}\Gamma_2$ and hence we can write $\alpha'_{21} = \alpha'_{11}(\Gamma_1^{-1}\Gamma_2)$ (i.e., as a function of α_{11} and the observable matrix Γ). Hence, the (structural) impulse vectors to shocks $1, \ldots, m_z$ satisfy:

$$\alpha^{[1]} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \begin{bmatrix} I_{m_z} \\ (\Gamma_1^{-1}\Gamma_2)' \end{bmatrix} \alpha_{11}, \tag{3.6}$$

where Γ_2 is $m_z \times (m - m_z)$ and Γ_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 α_{11}. It therefore restricts $(m - m_z)m_z$ elements of A.

The VAR implies further restrictions on α_{11}. To see this, it is useful to restate the VAR identification problem. From the covariance Σ of forecast errors alone, A is only identified up to multiplication by an orthonormal rotation matrix Q'. Denote the lower Cholesky factor of Σ by $\tilde{A}$. Then it holds true that $A = \tilde{A}Q$ (Uhlig, 2005, e.g.). Here, this implies that:

$$\left[\alpha^{[1]}, \alpha^{[2]}\right] = \tilde{A}Q = \left[\tilde{A}q^{[1]}, \tilde{A}q^{[2]}\right] \qquad \Rightarrow \alpha^{[1]} = \tilde{A}q^{[1]}$$

The fact that Q is a rotation matrix provides additional restrictions. In particular, it holds that:

$$I_{m_z} = (q^{[1]})'q^{[1]} = \alpha'_{11} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix}' \Sigma^{-1} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11} \quad \Leftrightarrow \quad \alpha_{11}\alpha'_{11} = \left(\begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix}' \Sigma^{-1} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \right)^{-1} \tag{3.7}$$

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

The following Lemma summarizes the above results:

Lemma 1. *(Stock and Watson, 2012; Mertens and Ravn, 2013) Under Assumption 1, the impact of shocks with narrative instruments is generally identified up to an $m_z \times m_z$ scale matrix α_{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 (3.6). Proof: See Appendix A.1.*

Thus, for $m_z > 1$, the extra data by themselves only identify a set of IRFs, and the statistical fit of the model is invariant to the particular choice of α_{11}.[7] To uniquely characterize impulse responses or historical shocks, however, I need further assumptions to partially identify specific responses. One simple assumption is that of a Cholesky structure. While not compelling in general,[8] I show in Section 3.2 that this is the appropriate choice for a class of DSGE models below. I therefore use this particular choice as my baseline but also provide an alternative statistical factorization.

Let me summarize the baseline factorization that follows Mertens and Ravn (2013) first. It is a Cholesky decomposition in a population two-stage least squares (2SLS) representation of previous the problem (i.e., given Σ, Γ). 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 $S_1S'_1$ and propose either an upper or a lower Cholesky

[7]Without extra structure, the model is thus only set identified in the sense of Moon and Schorfheide (2012).

[8]The countable number of Cholesky orderings does not span the uncountably many possible shock rotations consistent with the data when there are multiple shocks.

decomposition of S_1S_1'. To see this result mathematically, note that $A^{-1}v_t$ can be rewritten as the following system of simultaneous equations:

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \eta \\ \kappa & \mathbf{0} \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} + \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & S_2 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{3.8}$$

where $\eta = \alpha_{12}\alpha_{22}^{-1}$ and $\kappa = \alpha_{21}\alpha_{11}^{-1}$ are functions of Σ, Γ, given in Appendix A.2. Using the definition of $v_t = A\epsilon_t$ and simple substitution allows me to rewrite 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 Cholesky factorization as:

$$\alpha^{[1]} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \begin{bmatrix} (I-\eta\kappa)^{-1} \\ (I-\kappa\eta)^{-1}\kappa \end{bmatrix} \text{chol}(S_1S_1'). \tag{3.9}$$

The second factorization I consider is a generalization of the Cholesky decomposition of $\alpha_{11}\alpha_{11}'$, the contemporaneous forecast error variance attributable to the instrument-identified shocks $\epsilon_t^{[1]}$. The (lower) Cholesky 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.[9] The procedure described by Uhlig (2003) generalizes this by identifying the first instrument-identified shock as the shock that explains the most conditional forecast error variance in variable i over some horizon $\{\underline{h}, \ldots, \overline{h}\}$.[10]

With either factorization scheme, both the IRFs to the m_z policy shocks and the time series of the m_z policy shocks are identified given parameters. In my application, I look at both IRFs and historical shocks to explain different discrepancies in model fit. Next, I consider inference when the covariance matrix underlying the identification scheme has to be estimated.

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

[10] Following the steps in Uhlig (2003), this amounts to solving the following principal components problem:

$$\max_{\lambda_\ell} S\tilde{q}_\ell^\alpha = \lambda_\ell \tilde{q}_\ell, \quad \ell \in \{1,\ldots,m_z\}, \quad S = \sum_{h=0}^{\bar{h}} (\overline{h} + 1 - \max\{\underline{h}, h\}) \left(B^h \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \tilde{\alpha}_{11} \right)' e_1 e_1' \left(B^h \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \tilde{\alpha}_{11} \right).$$

Here $\alpha\tilde{}_{11}$ is the Cholesky factorization associated with (3.7), and e_1 is a selection vector with zeros in all except the first position. The desired α_{11} is given by $\alpha_{11} = \tilde{\alpha}_{11}q_\ell^\alpha$, where the eigenvectors $\tilde{q}_\ell^\alpha$ can be normalized to form an orthogonal matrix because S is symmetric and such that the signs of diag(α_{11}) are positive. Here I used that the variance whose forecast error variance is of interest is among the first m_z variables. Otherwise, redefine α_{11} to include κ or reorder variables as long as α_{11} is guaranteed to be invertible.

3.1.2 Posterior uncertainty

Here, I consider the case when the posterior over Γ is nondegenerate.[11] 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:

$$Y_{SUR} = X_{SUR}\beta_{SUR} + v_{SUR}, \qquad v_{SUR} \sim \mathcal{N}(0, V \otimes I_T), \tag{3.10}$$

where $Y_{SUR} = [\text{vec}(Y)', \text{vec}(Z)']'$ and v_{SUR} is defined analogously. In addition, I use the following definitions:

$$V = \begin{bmatrix} AA' & \Gamma' \\ \Gamma & \tilde{\Omega} \end{bmatrix} \qquad \beta_{SUR} = \begin{bmatrix} \text{vec}(B) \\ \text{vec}(\mu_z) \end{bmatrix},$$

$$X_{SUR} = \begin{bmatrix} I_{m_y} \otimes X_y & \mathbf{0}_{T(m_y p+1)\times Tm_z} \\ \mathbf{0}_{Tm_z \times Tm_y} & I_{m_z} \otimes X_z \end{bmatrix} \qquad X_y = \begin{bmatrix} Y_{-1} & \dots & Y_{-p} & \mathbf{1}_T \end{bmatrix}, \quad X_z = \begin{bmatrix} \mathbf{1}_T \end{bmatrix}.$$

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

$$\tilde{X} = ((U^{-1})' \otimes I_T) \begin{bmatrix} I_{m_y} \otimes X_y & \mathbf{0}_{T(m_y p+1)\times Tm_z} \\ \mathbf{0}_{Tm_z \times Tm_y} & I_{m_z} \otimes X_z \end{bmatrix} \qquad \tilde{Y} = ((U^{-1})' \otimes I_T) \begin{bmatrix} I_{m_y} \otimes Y & \mathbf{0}_{T(m_y p+1)\times Tm_z} \\ \mathbf{0}_{Tm_z \times Tm_y} & I_{m_z} \otimes Z \end{bmatrix}$$

$$N_{XX}(V) = \tilde{X}'\tilde{X} \qquad N_{XY}(V) = \tilde{X}'\tilde{Y}$$

$$S_T(\beta) = \tfrac{1}{\nu_0+T} \begin{bmatrix} (Y - XB)' \\ (Z - \mathbf{1}_T\mu_z')' \end{bmatrix} \begin{bmatrix} (Y - XB) & (Z - \mathbf{1}_T\mu_z') \end{bmatrix} + \tfrac{\nu_0}{\nu_0+T} S_0.$$

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

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

$$\bar{\beta}_T(V) = (N_{XX}(V) + N_0)^{-1}(N_{XY}(V) + N_0\bar{\beta}_0)$$

$$\beta | V, Y^T \sim \mathcal{N}(\bar{\beta}_T(V), (N_{XX}(V) + N_0)^{-1}), \tag{3.11a}$$

$$V^{-1} | \beta, Y^T \sim \mathcal{W}(S_T(\beta)^{-1}/(\nu_0 + T), \nu_0 + T). \tag{3.11b}$$

In general, no closed form posterior is available. The exception occurs when SUR collapses to ordinary least squares (OLS): 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

[11] When the prior and hence the posterior is close to a point mass, a metropolized sampling algorithm that exploits the fact that B, V are pinned down by θ may be more efficient. In the main application, I impose a relatively tight prior on the relevance of the instrument and thus also abstract from potentially weak instruments, an issue surveyed by Lopes and Polson (2014).

analogously for N_{XY}. In this special case, closed forms are available for the marginal distribution of V, allowing me to draw directly from the posterior. In general, however, the block closed form structure gives rise to a natural Gibbs sampler as in the following algorithm:

1. Initialize $V^{(0)} = S_T(\bar{\beta}_T)$.
2. Repeat for $i = 1, \dots n_G$:
   (a) Draw $\beta^{(i)}|V^{(i-1)}$ from (3.11a).
   (b) Draw $V^{(i)}|\beta^{(i)}$ from (3.11b).

3.2 Narrative DSGE-VAR

Having established identification and estimation in the purely narrative VAR, I now show that the conditional Cholesky method in the narrative BVAR recovers policy rule coefficients in a class of DSGE models. This class of models is invertible and only has Taylor-type policy rules, two conditions I define first. Then I adapt the idea of dummy variables (e.g., Del Negro and Schorfheide, 2004) to the previous SUR framework for estimating the narrative VAR.[12,13]

3.2.1 Invertibility

A necessary condition for the VAR and DSGE models to agree on the shocks affecting the economy is for both models to be able to span the same economic shocks. Fernandez-Villaverde et al. (2007) provide succinct sufficient conditions to guarantee that the economic shocks in the state space system (2.1) match up with those from the VAR (2.2):

Assumption 2. *A^* is nonsingular, and the matrix $C^* - D^*(A^*)^{-1}B^*$ is stable.*

Under this condition, it follows that the forecast errors from the VAR and the DSGE model coincide, as summarized in the following lemma:[14]

Lemma 3. *(Fernandez-Villaverde et al., 2007, p. 1022) Let Y_t be generated by the DSGE economy (2.1). Under Assumption 2, the variance-covariance matrix of the one-step-ahead prediction error in the Wold representation of Y_t is given by $\Sigma^* = (A^*)(A^*)'$.*

[12] For a given DSGE model, one can also test the identifying Assumption 1 parametrically by comparing a restricted model satisfying the Assumption with a more general model that violates it. Appendix C.2.1 reports the results of this exercise. In summary, I find evidence in favor of Assumption 1.

[13] 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.

[14] To see this, note that x_t can be expressed as a square-summable linear combination in terms of y^t. Hence, $\text{Var}[x_t|y^t] = 0$, and the Wold representation of y_t is given by:

$$y_t = B^* \sum_{j=0}^{\infty} (C^* - D^*(A^*)^{-1}B^*)^j D^*(A^*)^{-1} y_{t-1-j} + A^* \epsilon_t.$$

The one-step-ahead prediction error is, therefore, $y_t - \mathbb{E}[y_t|y^{t-1}] = A^*\epsilon_t$ with variance $(A^*)(A^*)'$.

I assume throughout that Assumption 2 holds so that a VAR(p) can approximate the DSGE model dynamics arbitrarily well. Thus, $AA' \approx A^*(A^*)'$. This assumption is not necessarily satisfied and, in general, depends on the observables Y_t. In my application with an equal number of AR(1) shock processes as observables, the only violation of Assumption 2 that I could find was a model with capital that did not include investment as an observable.[15] Figures C.12 and C.13 in the Appendix also show that the dynamics of the estimated VAR(4) approximation follow those of the pure DSGE model very closely.

3.2.2 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 do we 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 Cholesky compositions yield almost identical results. Here, I argue that, for a class of policy rules, the lower Cholesky decomposition actually recovers the true impact matrix $\alpha^{[1]}$ in the DSGE model.[16]

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 are 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.

Example 1. *An interest rate rule with observables only is a simple Taylor rule when output y_t and inflation π_t are observed:*

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

This rule maps into Definition 1 with $y_{p,t} = r_t$, $\lambda_p = e_r \rho_r + (1 - \rho_r)(-\gamma_y) e_y$, $\psi_1 = (1 - \rho_r)\gamma_\pi, \psi_2 = (1 - \rho_r)\gamma_y$, $\sigma_p = \omega_r$.

[15] I also verify this condition for each draw of the DSGE model parameters in my empirical application. Chari et al. (2005) also point to the challenge of recovering impulse responses in VAR models in economies with capital.

[16] Note that showing that two different Cholesky 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 as simple Taylor rules, however, the two Cholesky factorizations coincide, rationalizing the exercise in Mertens and Ravn (2013).

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

$$\begin{aligned} r_t &= (1-\rho_r)(\gamma_\pi \pi_t + \gamma_y \tilde{y}_t) + \rho_r r_{t-1} + \omega_r \epsilon_t^r \\ y_t^f &= B_{y^f}^* X_{t-1}^* + A_{y^f}^* \epsilon_t \end{aligned}$$

In general, the entire column vector $A_{y^f}^*$ *is nonzero, violating the exclusion restrictions for a simple Taylor rule.*

The key difference between examples 1 and 2 is that according to the rules in example 2, monetary policy also reflects other policy shocks contemporaneously despite controlling for output and inflation. But as long there is no more than general policy rule that violates Definition 1, the narrative VAR with Cholesky factorization still recovers the right IRFs.

Formally, the following proposition shows that in the DSGE model with simple policy rules, S_1 has a special structure that allows me to identify it uniquely using Γ, Σ, 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 (3.9).

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* $\Gamma = [G, \boldsymbol{0}]A^*$*. Then* $\alpha^{[1]}$ *defined in* (3.9) *satisfies* $\alpha^{[1]} = A^*[I_{m_z}, \boldsymbol{0}_{(m-m_z)\times(m-m_z)}]'$ *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 Cholesky decomposition in (3.9) works.

Formally, the Cholesky factorization of $S_1 S_1'$ proposed by Mertens and Ravn (2013) imposes the $\frac{m_z(m_z-1)}{2}$ 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 Cholesky structure, confirming that what is needed for identification via instrumental variables in the model is precisely the existence of $\frac{m_z(m_z-1)}{2}$ restrictions.[17] More generally, identification requires restrictions on the contemporaneous interaction between policy instruments that need not have the form of simple Taylor rules.

[17] In the notation of Appendix A.2, D_0 need only be lower triangular to achieve identification, not diagonal as implied

3.2.3 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 (β, V^{-1}), unless it collapses to a standard VAR model, the DSGE model implied prior also fails to generate unconditionally conjugate posteriors. Therefore, I consider an independent priors for the dynamics and the covariance matrix. In this approach, the prior is available in closed form, but the conditional prior variance of β is necessarily independent of V^{-1}, unlike in the standard DSGE-VAR model.

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

$$\bar{B}_0^y(\theta) = \mathbb{E}^{DSGE}[X_0 X_0' | \theta]^{-1} \mathbb{E}^{DSGE}[X_0 Y_0' | \theta] \quad \Leftrightarrow \quad \bar{\beta}_0^y(\theta) = \text{vec}(\bar{B}_0^y(\theta)) \tag{3.12a}$$

$$\bar{V}_0(\theta)^{-1} = \begin{bmatrix} A(\theta)^*(A(\theta)^*)' & (A(\theta)^*)[\text{diag}([c_1, \ldots, c_{m_z}]), \mathbf{0}]' \\ [\text{diag}([c_1, \ldots, c_{m_z}]), \mathbf{0}](A(\theta)^*)' & \omega_z^2 \, \text{diag}(A_1(\theta)^*(A_1(\theta)^*)') + A_1(\theta)^*(A_1(\theta)^*)' \end{bmatrix}, \tag{3.12b}$$

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 θ. $\bar{V}_0(\theta)$ satisfies Assumption 1 with $G = I_{m_z}$ and measurement error that is independent across instruments. The variance of the measurement error is parametrized to be ω_z^2 times the univariate shock variance.

The following dummy observations and likelihood implement the prior that $\beta \sim \mathcal{N}(\bar{\beta}_0, N_{XX}(\bar{V}_0))$ and $V^{-1} \sim \mathcal{W}(\bar{V}_0 T_0^V, T_0^V)$:

$$\text{vec}([Y_0^B, Z_0^B]) = \bar{X}_{0,SUR}(\theta)\bar{\beta}_0(\theta) + \mathbf{0}, \tag{3.13a}$$

$$\text{vec}([Y_0^B, Z_0^B]) \sim \mathcal{N}(\bar{X}_{0,SUR}(\theta)\bar{\beta}_0(\theta), \bar{V}_0(\theta) \otimes I_{T_0^B}) \tag{3.13b}$$

$$[Y_0^V, Z_0^V] = \mathbf{0} \times \beta + \bar{V}_0(\theta) \otimes I_{T_0^V}, \tag{3.13c}$$

$$\text{vec}([Y_0^V, Z_0^V]) \sim \mathcal{N}(\mathbf{0}, V \otimes I_{T_0^V}) \tag{3.13d}$$

where $X_{0,SUR}$ is the Cholesky factor of the following matrix:

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

in Proposition 1. Indeed, there are extra restrictions coming from the assumption of simple Taylor rules rather than allowing for a block-diagonal structure in the Taylor rules. Rather than imposing these overidentifying restrictions, I back out the implied Taylor-type policy rules and analyze the distribution of implied coefficient loadings.

Appendix C.3 provides additional details on the prior densities. The important distinction to a standard DSGE-VAR approach is that the coefficient prior depends on $\bar{V}_0(\theta)$ and not on the unknown covariance V.

The prior incorporates the Normal likelihoods 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 the corresponding marginal priors are given by:

$$\begin{aligned}
\beta|V^{-1},\theta &\sim \mathcal{N}(\bar{\beta}_0(\theta),(T_0^B \times \tilde{X}_0'(V^{-1}\otimes I)\tilde{X}_0')^{-1}),\\
\tilde{X}_0 &= \mathrm{diag}([X_0^y,\ldots,X_0^y,X_0^z,\ldots,X_0^z]),\\
V^{-1}|\beta,\theta &\sim \mathcal{W}_{m+m_z}(SSR_0(\beta,\theta)^{-1},T_0^V)\\
SSR_0(\beta,\theta) &= T_0^V \times \bar{V}_0(\theta) + T_0^B([Y_0(\theta),Z_0(\theta)]-X_0B(\beta))([Y_0(\theta),Z_0(\theta)]-X_0B(\beta))'.
\end{aligned}$$

For fixed θ (e.g., θ fixed at its prior mean $\bar{\theta}_0$) inference simply proceeds according to Lemma 2. To estimate θ, I need an extra step. Specifically, if θ has a nondegenerate prior distribution, I can simulate its posterior by sampling $\theta|\beta, V^{-1}$. Allowing for a nondegenerate distribution then also yields estimates of the structural parameters of the DSGE model as a byproduct of the DSGE-VAR estimation.

The prior and data density are given by:

$$\theta \sim \pi(\theta) \tag{3.14a}$$

$$\begin{aligned}
\pi(B,V^{-1}|\theta) &\propto |V^{-1}|^{-n_y/2}\ell(B,V^{-1}|Y_0(\theta),Z_0(\theta))\\
&= |V^{-1}|^{-n_y/2}f(Y_0(\theta),Z_0(\theta)|B,V^{-1})
\end{aligned} \tag{3.14b}$$

$$\tilde{f}(Y,Z|B,V^{-1},\theta) = f(Y,Z|B,V^{-1}). \tag{3.14c}$$

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

$$\begin{aligned}
\pi(\theta|B,V^{-1},Y,Z) &= \frac{f(Y,Z|B,V^{-1})\pi(B,V^{-1}|\theta)\pi(\theta)}{\int f(Y,Z|B,V^{-1})\pi(B,V^{-1}|\theta)\pi(\theta)d\theta} = \frac{p(B,V^{-1}|\theta)\pi(\theta)}{\int \pi(B,V^{-1}|\theta)\pi(\theta)d\theta}\\
&\propto \pi(B,V^{-1}|\theta)\pi(\theta),
\end{aligned} \tag{3.15}$$

as in the example in Geweke (2005, p. 77).

Adding a Metropolis-within-Gibbs step to the previous Gibbs sampler allows to simulate from the DSGE-VAR with a nondegenerate 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) Metropolis-Hastings within Gibbs:

(a) With probability p_h, set $i = h$, or else set $i = rw$.

(b) Draw a candidate θ_c from $\theta_c \sim \mathcal{N}(\theta_{(d-1)}), \Sigma)$. Assign unstable draws or draws violating the Fernandez-Villaverde et al. (2007) condition zero density.

(c) With probability $\alpha_{d-1,i}(\theta_c)$, set $\theta_{(d)} = \theta_c$, otherwise, set $\theta_{(d)} = \theta_{(d-1)}$.

$$\alpha_{d-1,i}(\theta_c) = \min\left\{1, \frac{\pi(B_{(d-1)}, V^{-1}_{(d-1)}|\theta_c)\pi(\theta_c)}{\pi(B_{(d-1)}, V^{-1}_{(d-1)}|\theta_{(d-1)})\pi(\theta_{(d-1)})} \frac{f_i(\theta_{(d-1)}|\theta_c)}{f_i(\theta_c|\theta_{(d-1)})}\right\}. \tag{3.16}$$

(d) Draw $B_{(d)}|\theta_{(d)}, V^{-1}_{(d-1)}$ given by (3.11a) and including dummy observations.

(e) Draw $V^{-1}_{(d)}|B_{(d)}, \theta_{(d)}$ given by (3.11b) and including dummy observations.

(f) If $d < D$, increase d by one and go back to (a), or 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.[18] To calibrate the covariance matrix of the proposal density, I use a first burn-in phase with a diagonal covariance matrix for the proposal density. The observed covariance matrix of the first stage is then used in subsequent stages up to scale. I use a second burn-in phase to calibrate the scale to yield an average acceptance rate across parameters and draws of 30%. To initialize the Markov chain, I then use a third burn-in phase whose draws are discarded. 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 by Chib and Ramamurthy (2010). Similar to their application to the Smets and Wouters (2007) model, I otherwise obtain a small effective sample size because of the high autocorrelation of draws when using a plain random-walk Metropolis-Hastings step.[19,20]

3.3 Marginal likelihood

The joint distribution of the data $[Y, Z]$, the VAR parameters β, V^{-1}, and the DSGE model parameters θ is given by

$$p(Y, Z, \beta, V^{-1}, \theta) = p(Y, Z|\beta, V^{-1})p(\beta, V^{-1}|\theta)p(\theta),$$

where equation (A.12) in the Appendix spells out the individual components.

[18] The t-distribution has 15 degrees of freedom as in Chib and Ramamurthy (2010).

[19] See also Herbst and Schorfheide (forthcoming).

[20] An alternative approach to increasing the efficiency of the sampler could be metropolization: Liu (1995) shows that, in a discrete setting, metropolization of the Gibbs sampler can lead to efficiency increases in terms of variance reduction. More generally, Robert and Casella (2005, ch. 10.3) discuss metropolization of the Gibbs sampler as a possibility to speed up the exploration of the parameter space. Here, metropolization is potentially interesting when the DSGE model prior becomes tight. Intuitively, a tight DSGE model prior means that movements in β become smaller. If the initial β is in an isolated region of the posterior, the resulting θ draws also tend to be drawn from the same region, and crossing over into other regions of the parameter space can be slow. This is not a problem in Del Negro and Schorfheide (2004) because they can draw directly from the marginal distribution of θ.

Integrating out the VAR and DSGE parameters gives the marginal likelihood:

$$p(Y,Z) = \int\int\int p(Y,Z,\beta,V^{-1},\theta)d\beta dV^{-1}d\theta \tag{3.17}$$

Since the prior is a function of T_0^V, T_0^B, the marginal likelihood is, implicitly, indexed by these prior hyperparameters. Next, I discuss the interpretation of the marginal likelihood in terms of these prior DSGE model weights. Then, I discuss the computation of the marginal likelihood.

3.3.1 Interpretation

Del Negro and Schorfheide (2004) introduce the prior DSGE model weight with the highest marginal likelihood as a diagnostic of misspecification: If the DSGE prior generates observations with properties like the data, these are informative and improve the model fit by reducing weight put on the wrong parameters. Del Negro et al. (2007) show in an AR(1) case with known variance that when the DSGE model and sample moments differ, there can be an interior optimum for the prior weight, trading off shrinkage with bias. In other cases, they find that the best-fitting prior weight can diverge, so that a high weight on the pure DSGE model or the (almost) flat prior VAR can emerge as optimal. Thus, the analysis of the prior weight that maximizes the marginal likelihood is meaningful and has a clear interpretation.

My model differs from Del Negro et al. (2007) in two dimensions: First, because of the extra information via instruments, my prior is not conjugate. Second, I allow for the weight on the covariance matrix and the dynamics to differ. I, therefore, characterize the behavior of the marginal likelihood in terms of the prior weight in the case of my model in Appendix A.8.1. For the case of known model dynamics, I characterize the marginal likelihood analytically and prove a lemma characterizing the slope of the marginal likelihood in the scalar case.[21] The appendix also provides a numerical example for the empirically relevant matrix case.

In summary, I show in Appendix A.8.1 that when the DSGE model prior fits the sample moments well, the marginal likelihood is strictly increasing in the number of dummy observations. In particular, the analytical results imply that the marginal likelihood is strictly increasing in T_0^V when the prior variance fits well enough. Vice versa, when the DSGE model fits the sample variance (sufficiently) poorly, the marginal likelihood is strictly decreasing in the prior weight T_0^V. A good fit of the model implied covariance matrix, therefore, shows in a higher optimal T_0^V that may even diverge to $+\infty$.

3.3.2 Computation

I combine the methods of Chib (1995) and Geweke (1999) to compute the likelihood (3.17).[22] Specifically, I first compute the marginal likelihood conditional on a specific θ_d – that is, the two inner integrals in (3.17) – using the method of Chib (1995) for models with fully conditional posteriors. This conditional marginal likelihood, when combined with the prior, gives the kernel of the θ posterior

[21] See Lemma 5 in Appendix A.8.1. In the analytic characterization, as in Del Negro et al. (2007), I consider the case of a single DSGE model parameter that maps directly into the VAR variance parameters.

[22] See Appendix A.5 for details.

that I use in the Geweke (1999) algorithm. While I find that relatively few draws – that is, 2,000 draws after 1,000 burn-in draws – are accurate to ± 0.1 log points, the repeated approximation takes time. Since the posterior draws $\{\theta_d\}_d$ are autocorrelated, I subsample every jth draw to yield a more efficient sample of 1,000 posterior draws that economizes on computing time.[23]

4 Empirical specification

4.1 Data and sample period

I use seven variables in the estimation. Importantly, I include information on the fiscal and monetary instruments: government spending, the average labor tax rate, and the effective Federal Funds Rate (FFR). In addition, I include standard macroeconomic variables: real GDP, real investment (including consumer durables), the debt-to-GDP ratio, and GDP inflation. GDP and its components are in per capita terms. Appendix C describes the details of the data construction, which follows Fernandez-Villaverde et al. (2015) for the tax time series. All fiscal variables consolidate the federal government with state and local governments.

To ensure my sample period covers both periods of significant variation in the fiscal variables, I start the estimation in 1947:Q1. This period includes Korean War expenditures as well as periods of declining and rising debt-to-GDP ratios. This can be important because Bohn (1991) has argued that long samples can be important since debt is slow moving. I stop the estimation in 2007:Q4, before the zero lower bound became binding. The resultant sample period is similar to that in Christiano et al. (2011a). Following Francis and Ramey (2009) and Ramey (2011), I allow for a quadratic trend.[24]

Narrative shock measures are taken from Romer and Romer (2004) for monetary policy shocks,[25] 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 the shocks in Romer and Romer (2010) that are not considered to be motivated by economic conditions.[26] In unreported robustness checks, I also consider alternative instruments for monetary policy shocks. First, I consider shocks identified by Kuttner (2001) and updated by Gürkaynak et al. (2005) based on federal funds futures changes around Federal Open Market Committee announcements. Second, I apply the same identification strategy to 13-week T-Bill futures

[23] I verify the pairwise rankings using the Meng and Wong (1996) bridge sampler with the iterative bridge function, as well as with plain importance sampling as described by Geweke (2005, ch. 8.2). Using these alternative methods, I confirm the shape of the marginal likelihood as a function of T_0, even though small differences exist between the pairwise Bayes factors. See Table A.1 in the Appendix. Another alternative would be to specify an explicit prior over $\tau = ({T_0}^B, {T_0}^V)$ and estimate τ as Adjemian et al. (2008) do for a standard DSGE-VAR. The advantage of my approach is that by only providing Bayes factors, I can report model fit independent of the prior over the hyperparameters τ.

[24] I detrend prior to estimation to match the detrended data with my stationary DSGE model.

[25] I also experimented with an update of their shock measures using the updated Greenbook data, available at `www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/philadelphia-data-set`. While the preliminary results are very similar, a potential issue is that toward the end of the Romer and Romer (2004) sample, there was a switch to chain weighting in the calculation of GDP that leads to persistent discrepancies between the underlying Greenbook data published on the `www.philadelphiafed.org` website and the original Romer and Romer (2004) data.

[26] Since instrument data are scarce, I follow Mertens and Ravn (2013) and set missing observations for instruments to zero. Alternatively, one could use data augmentation at the cost of including an extra step in the Gibbs sampler. We explore this possibility in Amir-Ahmadi et al. (2016).

that started trading more than a decade earlier, combining the announcement dates from Romer and Romer (2004) with those in Gürkaynak et al. (2005).[27]

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 by 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 formation. I add labor, capital, and consumption taxes as in Drautzburg and Uhlig (2015) and fiscal rules as in Leeper et al. (2010) and Fernandez-Villaverde et al. (2015). Here, I only discuss the specification of fiscal and monetary policy. The remaining model equations are detailed in Appendix A.6.

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

$$\hat{r}_t = \rho_r \hat{r}_{t-1} + (1-\rho_r)\big(\psi_{r,\pi}\hat{\pi}_t + \psi_{r,y}\tilde{y}_t + \psi_{r,\Delta y}\Delta\tilde{y}_t\big) + \xi_t^r, \tag{4.1}$$

where ρ_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. $\tilde{y}$ denotes the output gap (i.e., the deviation of output from output in a frictionless world). ξ_t^r follows an AR(1) process.[28]

The fiscal rules allow for both stabilization of output and the debt burden as well as smoothing of the different fiscal instruments:[29]

$$\hat{g}_t = \rho_g \hat{g}_{t-1} + (1-\rho_g)\left(-\psi_{g,y}\hat{y}_t - \psi_{g,b}\frac{\bar{b}}{\gamma\bar{y}}\hat{b}_t\right) + \xi_t^g \tag{4.2a}$$

$$\hat{s}_t = \rho_s \hat{s}_{t-1} + (1-\rho_s)\left(-\psi_{s,y}\hat{y}_t - \psi_{s,b}\frac{\bar{b}}{\gamma\bar{y}}\hat{b}_t\right) + \xi_t^s \tag{4.2b}$$

$$\frac{\bar{w}\bar{n}}{\bar{y}}d\tau_t^n = \rho_\tau \frac{\bar{w}\bar{n}}{\bar{y}}d\tau_{t-1}^n + (1-\rho_\tau)\left(\psi_{\tau^n,y}\hat{y}_t + \psi_{\tau^n,b}\frac{\bar{b}}{\gamma\bar{y}}\hat{b}_t\right) + \xi_t^{\tau,n} \tag{4.2c}$$

The disturbances ξ_t° follow exogenous AR(1) processes: $\xi_t^\circ = \rho_\circ \xi_{t-1}^\circ + \epsilon_t^\circ$.[30] Note that the sign of the coefficients in the expenditure components g_t, s_t is flipped so that positive estimates always imply consolidation in good times ($\psi_{\circ,y} > 0$) or when debt is high ($\psi_{\circ,b} > 0$).

[27] The source of the daily T-Bill data is Thomson Reuters' "CME-90 DAY US T-BILL CONTINUOUS" series from June 1976 to September 2003.

[28] Money supply is assumed to adjust to implement the interest rate and fiscal transfers are adjusted to accommodate monetary policy.

[29] Note that Leeper et al. (2010) assume there is no lag in the right-hand side variables, while Fernandez-Villaverde et al. (2015) use a one quarter lag.

[30] 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:[31]

$$
\begin{aligned}
&\frac{\bar{b}}{\bar{r}\bar{y}}(\hat{b}_t - \hat{r}) + \frac{\bar{w}\bar{n}}{\bar{y}}(d\tau_t^n + \bar{\tau}^n(\hat{w}_t + \hat{n}_t)) + \frac{\bar{c}}{\bar{y}}\bar{\tau}^c\hat{c}_t + \frac{(\bar{r}^k - \delta)\bar{k}}{\bar{y}}(d\tau_t^k + \bar{\tau}^k(\hat{r}_t^k\frac{\bar{r}^k}{\bar{r}^k - \delta} + \hat{k}_{t-1}^p)) \\
=&\hat{g}_t + \hat{s}_t + \frac{\bar{b}}{\gamma\bar{y}}(\hat{b}_{t-1} - \hat{\pi}_t)
\end{aligned} \tag{4.3}
$$

Following Christiano et al. (2011a), I include a cost channel of monetary policy: Firms have to borrow at the nominal interest rate to pay the wage bill at the beginning of the period. This allows monetary policy to cause inflation in the short run.[32]

The debate about variable selection in singular DSGE models is still ongoing; see Guerron-Quintana (2010), 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. Additionally, I include a total factor productivity (TFP) and investment-specific technology shock to explain GDP and investment, a price markup shock to contribute to inflation, and a shock to lump-sum transfers to add variation to the debt-to-GDP ratio.

Backing out the DSGE model implied historical shocks generally requires Kalman smoothing. Here, I exploit that under the invertibility Assumption 2, Lemma 3 implies that the data eventually fully reveal the hidden state variables of the DSGE model so that the contemporaneous uncertainty matrix for the state is zero. I initialize the Kalman filter with this matrix so that the Kalman smoother coincides with the Kalman filter.[33] I use Dynare (Adjemian et al., 2011) to solve the DSGE model for a given set of parameters.

Table 1: Calibrated parameters

Parameter	Value
Elasticity of substitution (inverse) σ	1.500
Discount rate (quarterly)	0.5%
Capital share α	0.300
Depreciation rate δ	0.020
Net TFP growth (quarterly)	0.4%
Steady state gross wage markup	1.150
Kimball parameters	10.000
Steady state government spending	0.200
Steady state consumption tax rate	0.073
Steady state capital tax rate	0.293
Steady state labor tax rate	0.165

To limit the dimensionality of the estimation problem, I calibrate a number of structural parameters and focus on the estimation of policy rules and shock processes (Table 1). These parameters largely

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

[32]This feature increases the marginal likelihood by about three log points in the benchmark specification.

[33]Since convergence to the fixed point is fast, the results are virtually unchanged when I initialize the filter with the stationary variance instead and use the Koopman (1993) Kalman smoother algorithm.

correspond to the prior mean in Smets and Wouters (2007). Average tax rates are calibrated as in Drautzburg and Uhlig (2015).

Priors for policy rules and shock processes are standard: I follow Smets and Wouters (2007) for common parameters and choose similar priors for new parameters of the fiscal rules. To parametrize the observation equations (3.2) of the instruments, I assume that both the matrix of loadings G and the covariance matrix of measurement errors are diagonal. My prior for the loadings c_i is that they are centered around unity, given the appropriate scaling of the narrative variables. Both the loadings and the relative standard errors of the measurement error have inverse gamma priors, so that the instruments are relevant for all parameter draws. The prior for the relative standard deviation is a relatively tight prior with a mean of 0.5 and a standard deviation of 0.1. This prior is intended to make the instruments informative to allow them to influence other parameter estimates: Overall the prior mean implies a signal-to-noise ratio of two in terms of standard deviations. Table C.2 in the Appendix lists all estimated parameters alongside their prior distributions.[34]

4.3 DSGE-VAR model specification

I use a VAR with $p = 4$ lags throughout this paper, estimated with the seven variables and three narrative instruments listed earlier. The specification of the lags follows Ramey (2011), but I also find that this finite lag approximation to the underlying DSGE model does well empirically: Figures C.13 and C.12 in the Appendix show that the VAR dynamics match that of the underlying estimated DSGE model. When I estimate the model with the DSGE model prior, I verify that the invertibility condition in Fernandez-Villaverde et al. (2007) holds for every draw.[35]

In the reported results, I do not control for exogenous variation in the instruments by including regressors X_t^z other than a constant. For robustness, however, I check if anticipated shocks or commodity prices drive the government spending or monetary policy instruments. This check is simple in the SUR framework: I add extra regressors that can soak up variation in the instruments. The regressors are the Fisher and Peters (2010) defense spending excess returns and the log-change in the producer price commodity index. I find no noticeable change in the results.

To calibrate the Gibbs sampler, I discard the first 50,000 draws as a burn-in period and keep every 20th draw thereafter until accumulating 5,000 draws. With a low prior weight on the DSGE model, this generates negligible autocorrelations of model summary statistics. The sampler is less efficient with a stronger DSGE model prior but performs well with the above sample size for moderate weights on the DSGE prior. Appendix C.3 also presents evidence on the convergence of the parameter estimates based on the Brooks and Gelman (1998) diagnostic.

[34]Increasing the prior standard deviation increases the marginal likelihood but otherwise leads to qualitatively similar results. In earlier stages, I tested the assumption that the matrix G is diagonal under full information, and I found evidence in favor of the current parsimonious parametrization. See Appendix C.2.1.

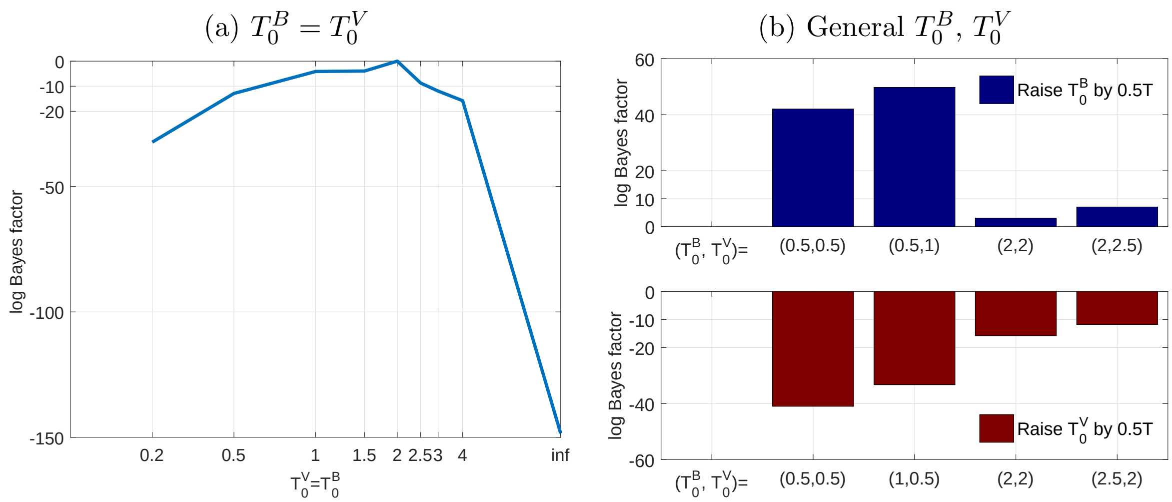
[35]Experimenting with the lag length in the flat prior VAR did not change the results quantitatively. Information criteria in a DSGE-VAR with the structural parameters fixed at their prior means indicated an optimal lag length of two or three.

5 Results

I choose to compare the estimated DSGE-VAR with its pure DSGE model counterpart based on the joint distribution of structural parameters. Alternatively, I could have computed shocks and IRFs under the VAR approximation to the DSGE model. However, the implied model dynamics are indistinguishable for almost all variables, and the discrepancies that do arise are small (see the estimated IRFs in Figures C.12 and C.13).

5.1 Marginal likelihood

Figure 1 shows the Bayes factors implied by the marginal likelihood. Panel (a) shows the likelihood when T_0^B moves in lockstep with T_0^V. The marginal likelihood initially increases but then falls first slightly and then sharply with the weight on the DSGE model. There is thus clear evidence for an interior peak as in Del Negro et al. (2007), here estimated at $T_0 = 2 \times T$.[36] Thus, a pure DSGE model is misspecified compared with the data.



The marginal narrative DSGE-VAR likelihood peaks at an interior overall weight on the DSGE model restrictions: On the grid, $T_0 = 2$ (i.e., adding two full samples worth of observations) yields the highest data density. Panel (b) shows Bayes factors comparing models when T_0^B and T_0^V differ. At the grid points I considered, increasing the weight on the dynamics strictly increases the fit, while increasing the prior weight on the DSGE model covariance structure significantly lowers the fit. These effects weaken as the overall prior weight increases. This is evidence against the DSGE model implied identification via its contemporaneous covariance structure.

Figure 1: Marginal likelihood for varying DSGE model weights

However, it is unclear whether the overall likelihood is driven by the (mis-)specification of the dynamics or the shock identification. This is explored in the right panel of Figure 1. It shows that the DSGE dynamics agree with the data to a significant degree: The likelihood is increasing when adding prior weight to the DSGE dynamics. In contrast, the likelihood is decreasing significantly when increasing the prior weight on the DSGE model covariance matrix. This suggests that the DSGE model

[36] In Table A.1, I provide a comparison with two alternative estimators for the likelihood that confirm the overall shape reported here.

can capture the sample dynamics reasonably well, but the implied shock identification is problematic. These qualitative results also hold for models estimated separately with either the fiscal or monetary shock proxies and for different priors (see Figures C.4 and C.5 in the Appendix).

To better understand the model fit along the two dimensions of shock identification and dynamics, I proceed by presenting the IRFs that shape the model dynamics. Subsequently, I turn to the policy shocks directly to gain intuition on the identification. Policy rule estimates and, more generally, DSGE model parameter estimates provide an additional way to understand the results summarized by the likelihood. For presenting the historical shocks and the estimated IRFs, I focus on the DSGE-VAR model with a weak prior to highlight the discrepancy between the sample and prior moments that determine the marginal likelihood.

5.2 IRFs

5.2.1 Benchmark DSGE-VAR estimates

Figure 2 shows the responses of output to the three policy shocks in the estimated DSGE-VAR: Shown in black and shades of gray are the pointwise posterior median and 68% and 90% credible sets. Start with the government spending shock. A 1% shock leads to an additional buildup in government spending that declines but persists for more than five years. Output rises persistently, too. The impact response of output is centered slightly below 0.4. This corresponds to an impact multiplier slightly below 2. While this value is large, the persistence of the spending increase implies that the wealth effects are also large.

Tax shocks lead to an initially hump-shaped and persistent drop in output. While the taxes decline smoothly back to their normal level after the initial rise, there is evidence that this reversion to zero may last more than five years and thereby explain the persistence in the output drop.

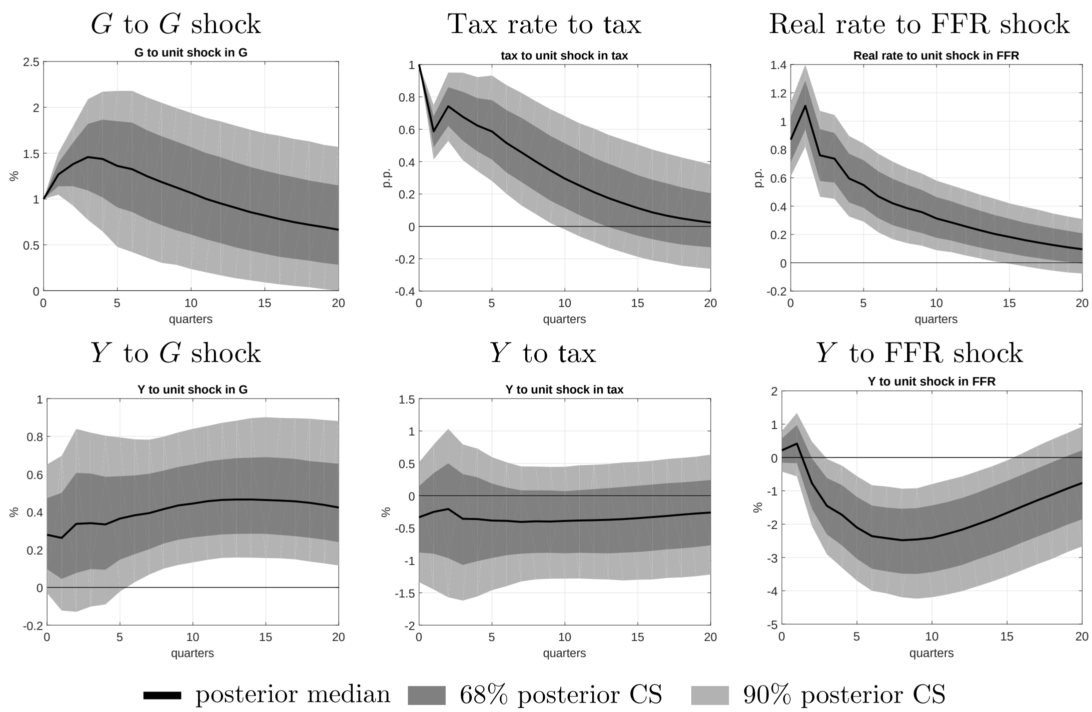
Shocks to monetary policy, in contrast to fiscal shocks, last only about two years. The rise in the nominal interest rate causes the real rate to rise and leads, with a delay of half a year, to a hump-shaped drop in output for up to two years after the shock. In addition, there is a double dip after four years. This double dip, as well as the large effects of fiscal policy, may be related to the fiscal policy interactions that I discuss later.

Besides the substantive lessons of the narrative DSGE-VAR, the wide credible sets stand out: For example, the 68% posterior credible set for the impact spending multiplier ranges from about 1.25 to 2.5. However, these wide bands represent an improvement over the even wider credible sets in the narrative VAR with a flat prior discussed next.[37]

5.2.2 Robustness: flat prior and different identification schemes

Here I briefly discuss how the DSGE model prior affects inference relative to a flat prior VAR and whether the conclusions are robust to departure from the conditional Cholesky factorization that

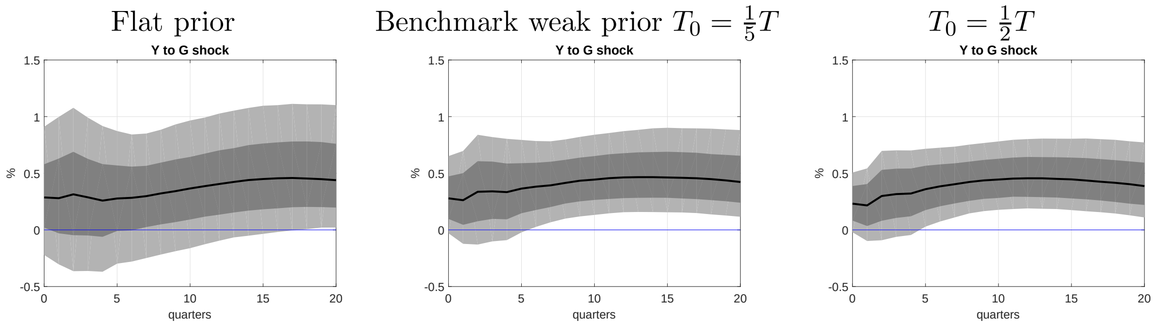
[37] Appendix B shows for the flat prior VAR with a single instrument that the SUR-based credible sets also have reasonable sampling properties.



There is a buildup in government spending in response to a government spending shock, causing a significant and lasting increase in output. A tax shock raises tax rates persistently but implies a smooth decline to zero. With a one quarter lag, output drops persistently. A FFR shock increases the real rate and causes a significant output drop after two quarters, with no significant response on impact.

Figure 2: VAR response with $T_0^B = T_0^V = \frac{1}{5}T$ DSGE model prior

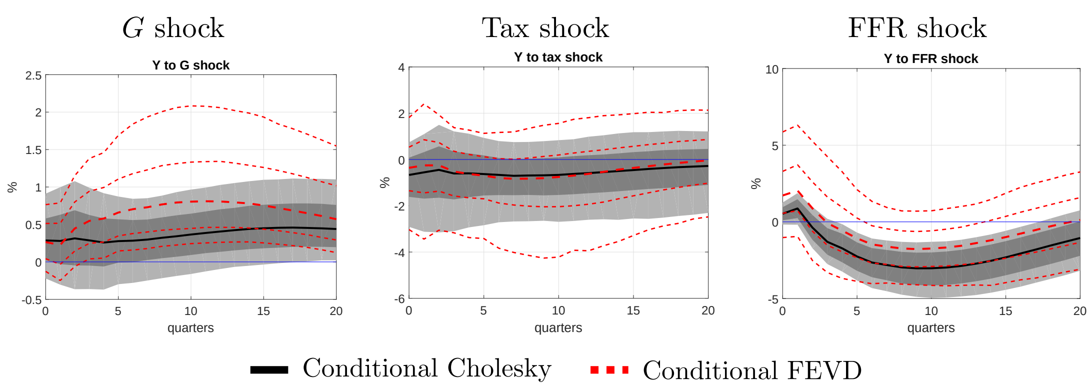
Taylor-type policy rules imply. Figure 3 shows that moderate amounts of prior information, such as adding data worth one fifth of the U.S. postwar sample, already sharpen the inference significantly.



A weak DSGE model prior tightens the posterior bands substantially, as illustrated here with the output response to a government spending shock.

Figure 3: Effects of the prior precision on the posterior uncertainty: response of output to government spending

Figure 4 takes the robustness check one step further. It compares two different factorizations of the jointly identified effects of the narrative shocks. While I have shown earlier that the conditional Cholesky factorization is justified in DSGE models with Taylor-type policy rules, one may be interested in what happens when we relax this assumption. In addition to the benchmark conditional Cholesky procedure, I also show the results from maximizing the conditional forecast error variance decomposition (FEVD). I find that the same qualitative features remain. Specifically, the persistent and uncertain output responses to either fiscal shock remain. Also, the delayed contraction to a monetary policy shock shows in both of the identification schemes, although the posterior uncertainty increases significantly for the impact response. These wider posterior bands are natural in the FEVD because it also depends on the uncertain VAR dynamics.



This figure shows two analogues to the output responses in the bottom panel of Figure 2 with a flat VAR prior and two different identification schemes. The output responses are qualitatively the same with a flat prior and the DSGE model prior but more uncertain. The qualitative results for the fiscal policy shocks are also largely unchanged when the shocks are decomposed by maximizing the forecast error variance (FEVD) due to the fiscal shocks over the first 20 quarters, except that the contraction of output in response to a monetary policy shock becomes less significant.

Figure 4: Robustness of output effects across identification schemes with a flat prior

5.2.3 DSGE-VAR vs DSGE model

A key component that enters the marginal likelihood is the particular model dynamics. This section discusses the shortcomings of the DSGE model relative to the DSGE-VAR impulse responses.[38] In particular, the pure DSGE model misses policy interactions and tax and inflation dynamics.

For the government spending shock in Figure 5,[39] the first order mismatch is that the DSGE model understates the speed of the government spending buildup. I, therefore, scale the government spending shock to yield the same median cumulative government spending over the first five years, that is, roughly matching its present discounted value. The second striking fact about the identified responses is that both inflation and the FFR drop, both in the DSGE-VAR and the pure DSGE model. This is possible even though monetary policy in the model pays attention only to output and inflation. However, the pure DSGE model understates also the initial increase in output, presumably because it predicts a drop in investment.

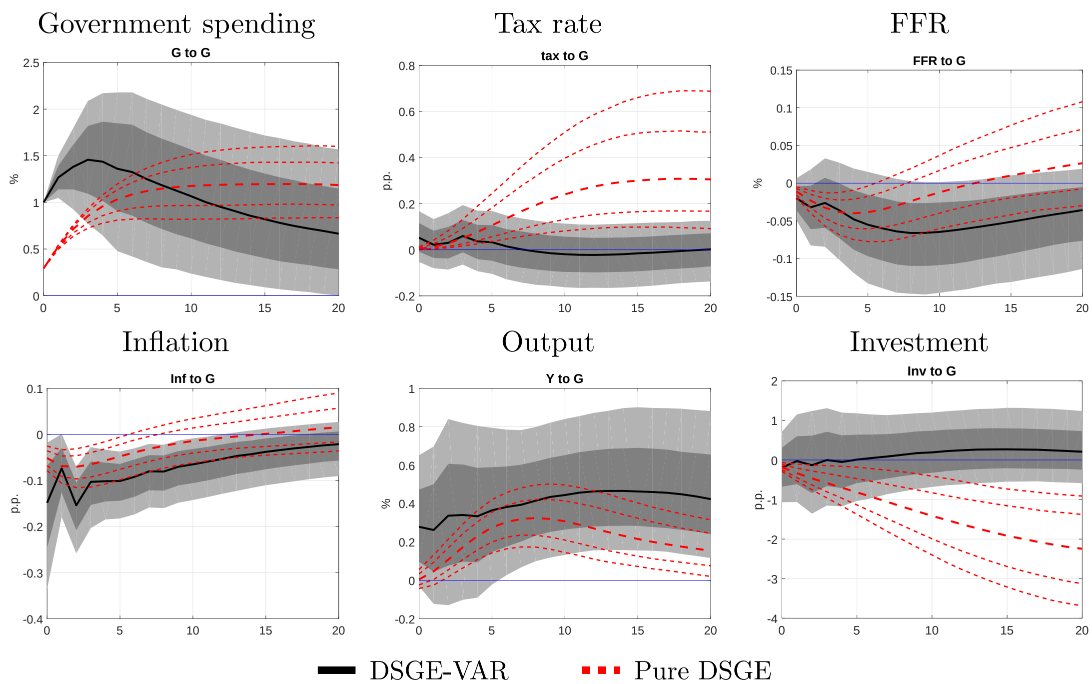


Figure 5: Full set of responses to spending policy shock with weak prior $T_0 = \frac{1}{5}T$

Figure 6 shows the responses to a tax rate shock. The responses reveal that the DSGE model struggles to match the dynamics of the tax rate and, hence, the inflation and nominal interest rate: The DSGE-VAR tax increases are less persistent and cause only short-lived increases in inflation and interest rates. Otherwise, the responses largely agree, not least because of the large error bounds in

[38] I choose to present results based on the state-space representation of the DSGE model. None of the results depend on this, since the DSGE-VAR and the pure DSGE model produce virtually identical responses. See Figures C.12 and C.13 in the Appendix.

[39] The responses of debt to all shocks are shown in Figure C.6 in the Appendix, immediately before the responses in the best-fitting DSGE-VAR.

the DSGE-VAR. The DSGE model explains the ambiguous output response to a tax increase with increased in government spending in response to a lower debt burden in the economy.

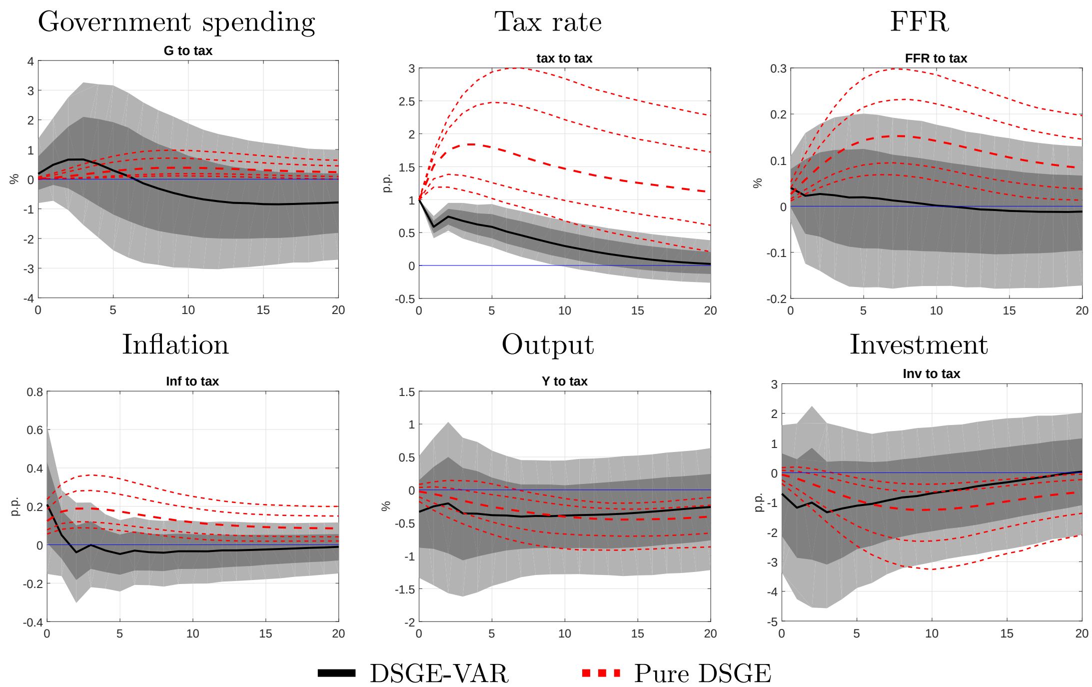


Figure 6: Full set of responses to tax rate shock with weak prior $T_0 = \frac{1}{5}T$

Are the estimated reactions of monetary policy to fiscal policy in the DSGE-VAR a plausible benchmark? Romer and Romer (2014) provide qualitative evidence that the Federal Reserve has indeed considered fiscal policy in its monetary policy decisions. They document staff presentations to the 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 during at least one FOMC meeting: According to Romer and Romer (2014), the Federal Reserve Bank 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 attempts to stimulate demand through fiscal transfer programs (p. 41f.). The DSGE-VAR similarly produces a lower interest rate in response to government spending increases and only short-lived responses to tax increases.

The responses to the monetary policy shock in Figure 7 are most precisely estimated and exhibit the largest discrepancies between the DSGE-VAR and the pure DSGE model. The latter struggles to explain the responses of output, investment, and inflation reminiscent of the price puzzle. This is despite including a cost channel of monetary policy as in Christiano et al. (2011a).[40] The DSGE model also fails to match the sizable fiscal contraction at the one-year horizon that the DSGE model implies.

[40] Controlling for commodity prices in the flat prior VAR does not affect these results. It may hint at another dimension of misspecification of monetary policy shock measures. For example, Caldara and Herbst (2015) argue that monetary policy also reacts to credit spread shocks.

While the model can match most dynamics in response to fiscal shocks with a tighter DSGE model prior, the model continues to have problems to match the estimated dynamics following a monetary policy shock. See Figures C.7 to C.9 in the Appendix.

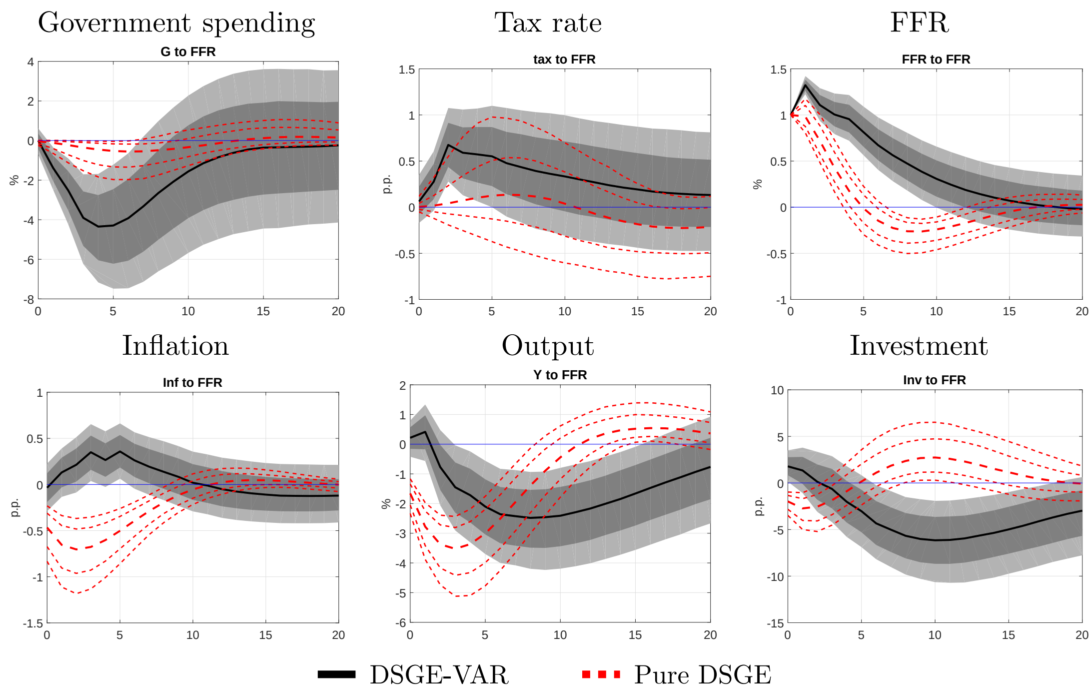


Figure 7: Full set of responses to FFR shock with weak prior $T_0 = \frac{1}{5}T$

5.3 Historical shocks

What is the relationship between the identified policy shocks in the DSGE-VAR, the DSGE model, and their narrative proxies? Importantly, the three shock proxies have median correlations between 0.5 and 0.65 with the corresponding identified shocks, with lower bounds of 0.37 or higher. Thus, the shock proxies are of reasonably good quality. See the middle column in Table 2(a).

The combined DSGE-VAR and the pure DSGE model also agree regarding the identified shocks across different subsamples, as Figure 8 exemplifies for the government spending shock.[41,42] The overall correlation between the DSGE-VAR spending shock and its DSGE counterpart is 0.62. The correlation for tax shock is slightly lower at 0.55, while the monetary policy shock matches up very well across the two models, with a median correlation of 0.82; see Table 2(a). As panel (b) of the same table shows, this correlation mechanically increases with the prior weight on the DSGE model.

Note that the pure DSGE has a notably lower correlation with the narrative shock proxies than

[41] Figures C.10 and C.11 in the Appendix show these subsamples for the tax and monetary policy shocks.

[42] Comparing shocks is in the spirit of the work by Rudebusch (1998). Sims (1998) disputes that comparing different historical monetary policy shocks across different VARs in Rudebusch (1998) is meaningful. Note that their dispute involves VARS with differing information sets, whereas I compare models with identical information sets except for the instruments. Since the instruments do not add information in the DSGE model, forecasts and forecast errors from these models should be close if the VAR approximation is accurate, as I assume in Proposition 1. Comparing the subset of identified shocks that I extract from these forecast errors is therefore a meaningful exercise.

Table 2: Historical shock comparison: full sample correlation

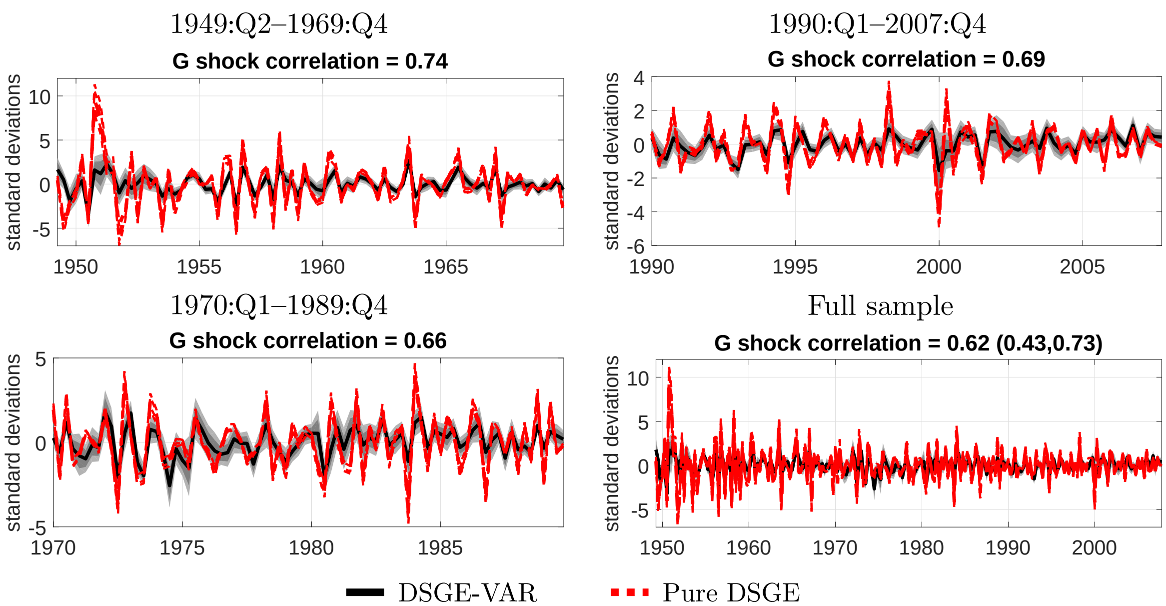
(a) Weak prior: $T_0 = \frac{1}{5}T$

	DSGE-VAR vs DSGE		DSGE-VAR vs IV		DSGE vs IV	
Shock	Median	(90% band)	Median	(90% band)	Median	(90% band)
G	0.62	(0.43, 0.73)	0.50	(0.37, 0.59)	0.40	(0.37, 0.43)
Tax	0.55	(0.37, 0.69)	0.56	(0.39, 0.69)	0.30	(0.19, 0.40)
FFR	0.82	(0.74, 0.88)	0.65	(0.63, 0.67)	0.58	(0.55, 0.60)

(b) DSGE-VAR vs DSGE

	$T_0 = \frac{1}{5}T$		$T_0 = 2 \times T$		$T_0 = 4 \times T$	
Shock	Median	(90% band)	Median	(90% band)	Median	(90% band)
G	0.62	(0.43, 0.73))	0.77	(0.70, 0.83)	0.81	(0.75, 0.86)
Tax	0.55	(0.37, 0.69)	0.73	(0.60, 0.82)	0.79	(0.67, 0.86)
FFR	0.82	(0.74, 0.88)	0.83	(0.75, 0.88)	0.87	(0.82, 0.91)

Shown are the posterior median correlations (and posterior credible sets) between the identified VAR shocks and the corresponding structural shocks from the DSGE model and their respective correlation with the nonzero instrumental variables (IV). Underlying are draws from the joint posterior of the structural parameters θ and the corresponding VAR parameters.



The solid lines represent the DSGE-VAR shocks; the dashed lines represent the corresponding DSGE model shocks. The plot of the entire sample shows that the shocks are reasonably close to *iid*. Their correlation is high and robust across subsamples.

Figure 8: Historical government spending shocks

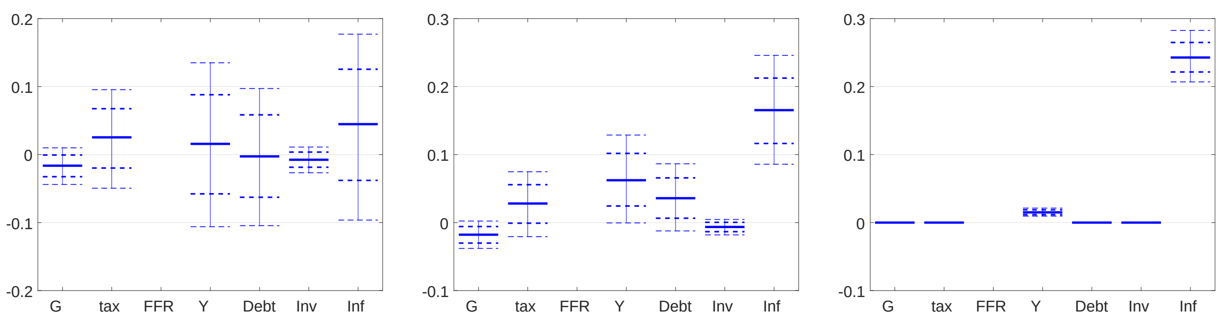
the DSGE-VAR. This indicates that the model may lack the flexibility to fit the narrative data and the standard time series simultaneously; see the upper panel of Table 2.

5.4 Policy rules

Partial identification of policy shocks implies identification of the underlying policy rules. Arias et al. (2015) exploit this, and it underlies the link between the narrative instruments and the DSGE model here. It is therefore instructive to back out the underlying policy rules to understand the mechanics of the DSGE-VAR model. The implied policy rules itself, however, should be interpreted with caution: Christiano et al. (1999), for example, show how an econometrician can recover the correct IRF to a monetary policy shock but estimate the wrong policy rule when the policymaker observes the data only imperfectly.

The raw data are rather uninformative about policy rules: The lower panel in Table 3 shows that the policy rule coefficients have wide posterior confidence bands that often include zero. This is where the DSGE model prior helps: It pulls the imprecise estimates of the policy rule coefficients toward those implied by the DSGE model prior. Figure 9 illustrates this clearly for the monetary policy rule: With the DSGE model prior, the policy rule estimates imply that the Fed increases interest rates in response to higher inflation or output growth.

Weak prior: $T_0 = \frac{1}{5}T$ | Best-fitting model: $T_0 = 2 \times T$ | Dogmatic prior: $T_0 = \infty$



Shown are the posterior median and 68% and 90% credible set of the monetary policy rule coefficients implied by the partially identified VAR with a flat prior and the DSGE prior. With a flat prior, there is no systematic pattern. With the DSGE-VAR prior, the model implies monetary tightening in response to higher inflation but also a reaction to fiscal policy.

Figure 9: Monetary policy rule estimates

5.5 Parameter estimates

As a byproduct of the estimation of the DSGE-VAR, the estimator also yields estimates of the structural DSGE model parameters. For brevity, I present the full estimates only in Table C.2 in the Appendix. As shown there, the tighter the DSGE model prior, the higher the estimated stickiness and adjustment costs.[43] With few exceptions, the posterior also differs significantly from the prior, and the

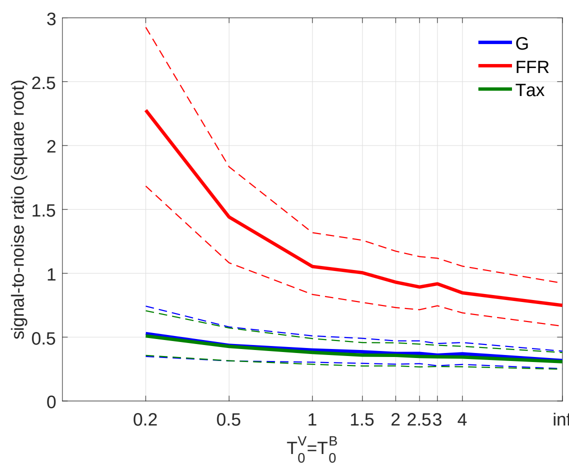
[43] Figure C.17 in the Appendix uses the Brooks and Gelman (1998) diagnostic to show that the posterior simulator for θ has converged reasonably well.

Table 3: Policy rule estimates

	G	tax	FFR	Y	Debt	Inv	Inf
Dogmatic prior: $T_0 = \infty$							
G		0	0	-0.05 (-0.12, 0.02)	-0.01 (-0.02, -0.00)	0	0
tax	0		0	0.02 (0.01, 0.03)	0.01 (0.00, 0.01)	0	0
FFR	0	0		0.02 (0.01, 0.02)	0	0	0.24 (0.21, 0.28)
Best-fitting model: $T_0 = 2 \times T$							
G		0	0	-0.45 (-1.74, 0.44)	-0.17 (-1.01, 0.49)	0.04 (-0.12, 0.24)	0.36 (-0.60, 1.34)
tax	0.01 (-0.08, 0.09)			-0.03 (-0.36, 0.35)	-0.04 (-0.29, 0.24)	0.00 (-0.06, 0.07)	-0.08 (-0.43, 0.28)
FFR	-0.02 (-0.04, 0.00)	0.03 (-0.02, 0.07)		0.06 (-0.00, 0.13)	0.04 (-0.01, 0.09)	-0.01 (-0.02, 0.00)	0.17 (0.09, 0.25)
Weak prior: $T_0 = \frac{1}{5}T$							
G		0	0	0.52 (-2.92, 2.82)	1.06 (-1.27, 3.33)	0.20 (-0.17, 0.90)	2.61 (-0.08, 7.20)
tax	0.01 (-0.25, 0.17)		0	-0.16 (-0.91, 1.01)	-0.22 (-0.84, 0.71)	-0.00 (-0.15, 0.15)	-0.43 (-1.44, 0.71)
FFR	-0.02 (-0.04, 0.01)	0.03 (-0.05, 0.10)		0.02 (-0.11, 0.14)	-0.00 (-0.10, 0.10)	-0.01 (-0.03, 0.01)	0.04 (-0.10, 0.18)

Shown are the posterior median and 90% credible set of the policy rule coefficients implied by the partially identified DSGE-VAR. The coefficients on policy interactions have a triangular pattern by construction. With a weak prior, the posterior over the coefficients is very dispersed. The best-fitting model implies monetary tightening in response to higher inflation and some indications of accommodating fiscal policy. The estimates for fiscal policy rules are very dispsersed and are bounded away from zero only with a very strong prior.

policy rule estimates are economically meaningful. The exception is the wage indexation parameter, which is plausible because I do not use wages in the estimation. I now focus on the parameters linking the DSGE model to the observed narrative shocks.



The plot shows the signal-to-noise ratio as a function of the prior weight given to the DSGE model. The ration is defined as the standard deviation of the instrument attributable to the structural shock divided by the standard deviation of the measurement noise. The prior signal-to-noise ratio in the DSGE model is 2. The plot implies that the data is informative about the signal-to-noise ratio that falls initially quickly and then stabilizes around 0.4 for the two fiscal instruments and falls toward 0.75 for the monetary policy shock.

Figure 10: Signal-to-noise ratio of instruments with varying DSGE model weight

The estimates include the parameters that link the narrative instruments and the structural shocks. Figure 10 plots the posterior of the implied signal-to-noise ratio for the three instruments as a function of the weight on the DSGE model. It shows that the posterior signal-to-noise ratio decreases initially quickly and then stabilizes around 0.4 for the two fiscal instruments and falls toward 0.75 for the monetary policy shock. This may indicate that the monetary policy shock instrument is measured more precisely than the fiscal shock instruments. This is in line with the evidence from the historical shocks that the narrative monetary policy shock proxies match up the best with the identified shocks.

6 Conclusion

A key question for academics and practitioners using quantitative DSGE models is whether these models agree with results derived from methods that, in the words of Sims (2005, p. 2), "get by with weak identification." This paper takes a step toward formally assessing the potential misspecification along the dimension of shock identification, extending earlier work (Del Negro et al., 2007) that leaves out the question of shock identification.

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 Taylor rules with limited direct interaction between policy instruments. Second, it shows how to estimate a narrative VAR using the standard SUR framework and how to incorporate a DSGE model prior.

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 improves the overall statistical fit of the DSGE-VAR model as measured by the marginal data density. However, by varying separately the prior weight on the DSGE model dynamics and covariance, I show that this fit comes from matching the dynamics rather than the covariance structure in the data. Looking at impulse responses shows that the best-fitting DSGE-VAR and the corresponding pure DSGE model largely agree on the dynamics following fiscal shocks. In contrast, the dynamics following a monetary policy shock differ. This may reflect the fact that the FOMC has already practiced forward guidance in monetary policy since prior to the Great Recession (see Campbell et al., 2012) and that the monetary policy shock measures also capture this news component. Looking at historical shocks, the pure DSGE fits the estimated fiscal shock worse than the monetary shocks, as well as their respective proxies.

Several extensions are possible. First, from a substantive perspective, it would be of interest to assess the degree of misspecification of different medium-scale models to isolate model ingredients that improve the DSGE models' fit. Here, for example, I introduce a cost channel from Christiano et al. (2011a) that improves the model fit slightly but does not overcome the estimated price puzzle. A more systematic way would be to use the Macroeconomic Model Database introduced in Wieland et al. (2012). Since their DSGE model code is standardized, my narrative DSGE-VAR framework can readily be used to compare a variety of different models.

Second, interesting methodological improvements are possible. Within the Bayesian paradigm, imputing missing values for the instruments simply amounts to an extra step in the Gibbs sampler. In Amir-Ahmadi et al. (2016), we are currently exploring this possibility while also allowing for stochastic volatility in a factor-augmented VAR (FAVAR). The FAVAR framework also allows us to address challenges arising from anticipated effects. Here, I addressed these challenges only by using control variables. While these extensions still use instruments for identification, it is also be possible to test identifying assumptions based only on stochastic volatility alone (e.g., Lanne et al., 2015), and this approach could also be applied to assess DSGE models.

What lessons go beyond my specific application? I find that narrative macro shocks alone allow only for imprecise shock identification. This suggests that there are limits to applying the theoretically work on instrument identification in VARs more broadly unless we can systematically generate additional information. In this paper, economic theory generates this information through the DSGE model prior. In ongoing work in a panel VAR, Carlino and Drautzburg (2015) build instruments based on cross-sectional heterogeneity in regional micro data. Both micro data and the combination of economic theory to generate priors seem promising for instrument-based shock identification in macro models.

References

Adjemian, Stephane, Houtan Bastani, Michel Juillard, Ferhat Mihoubi, George Perendia, Marco Ratto, and Sebastien Villemot, "Dynare: Reference Manual, Version 4," Dynare Working Papers 1, CEPREMAP 2011.

—, Matthieu Darracq Paries, and Frank Smets, "A Qquantitative Perspective on Optimal Monetary Policy Cooperation Between the US and the Euro Area," Working Paper Series 0884, European Central Bank March 2008.

Amir-Ahmadi, Pooyan, Thorsten Drautzburg, and Christian Matthes, "Evolving Dynamics in Monetary-Fiscal Interaction: Fundamentalness and Identification in a High-Dimensional Environment," 2016. FRB Philadelphia, draft.

Arias, Jonas E., Dario Caldara, and Juan Rubio-Ramirez, "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure," 2015. unpublished, Duke University.

Blanchard, Olivier, and Roberto Perotti, "An Empirical Characterization of The Dynamic Effects of Changes in Government Spending and Taxes on Output," *The Quarterly Journal of Economics*, November 2002, *117* (4), 1329–1368.

Bohn, Henning, "Budget balance through revenue or spending adjustments?: Some historical evidence for the United States," *Journal of Monetary Economics*, 1991, *27* (3), 333–359.

Brooks, Stephen P., and Andrew Gelman, "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, 1998, pp. 434–455.

Caldara, Dario, and Edward Herbst, "Bayesian Proxy SVAR: Theory and Application to Monetary Policy," 2015, unpublished. Federal Reserve Board of Governors.

Campbell, Jeffrey R., Charles L. Evans, Jonas D.M. Fisher, and Alejandro Justiniano, "Macroeconomic Effects of Federal Reserve Forward Guidance," *Brookings Papers on Economic Activity*, 2012, *44* (1 Spring), 1–80.

Canova, Fabio, Filippo Ferroni, and Christian Matthes, "Choosing the Variables to Estimate Singular DSGE Models," CEPR Discussion Papers 9381, March 2013.

—, —, and —, "Approximating Time Varying Structural Models With Time Invariant Structures," 2015. unpublished, Federal Reserve Bank of Richmond.

Carlino, Gerald, and Thorsten Drautzburg, "All Jobs Are Created Equal? Examining the Importance of Startups for Local Labor Demand," 2015. unpublished, FRB Philadelphia.

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan, "A Critique of Structural VARs Using Real Business Cycle Theory," Technical Report 2005.

Chib, Siddhartha, "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 1995, pp. 1313–1321.

— and Srikanth Ramamurthy, "Tailored Randomized Block MCMC Methods with Application to DSGE Models," *Journal of Econometrics*, 2010, *155* (1), 19–38.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, "Monetary Policy Shocks: What Have We Learned and to What End?," in J. B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, Vol. 1 of *Handbook of Macroeconomics*, Elsevier, September 1999, chapter 2, 65–148.

_ , _ , and _ , "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, February 2005, *113* (1), 1–45.

_ , Mathias Trabandt, and Karl Walentin, "DSGE Models for Monetary Policy Analysis," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3 of *Handbook of Monetary Economics*, Elsevier, 2011, chapter 7, 285–367.

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo, "When Is the Government Spending Multiplier Large?," *Journal of Political Economy*, 2011, *119* (1), 78 – 121.

Clarida, Richard, Jordi Galí, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, February 2000, *115* (1), 147–180.

Del Negro, Marco, and Frank Schorfheide, "Priors from General Equilibrium Models for Vars," *International Economic Review*, 2004, *45* (2), 643–673.

_ , _ , Frank Smets, and Rafael Wouters, "On the Fit of New Keynesian Models," *Journal of Business & Economic Statistics*, April 2007, *25*, 123–143.

Drautzburg, Thorsten, and Harald Uhlig, "Fiscal Stimulus and Distortionary Taxation," *Review of Economic Dynamics*, October 2015, *18* (4), 894–920.

Faust, Jon, "The New Macro Models: Washing Our Hands and Watching for Icebergs," *Sveriges Riksbank Economic Review*, 2009, 45–68.

Fernandez-Villaverde, Jesus, Juan F. Rubio-Ramirez, Thomas J. Sargent, and Mark W. Watson, "ABCs (and Ds) of Understanding VARs," *American Economic Review*, June 2007, *97* (3), 1021–1026.

_ , Pablo A. Guerron-Quintana, Keith Kuester, and Juan Rubio-Ramirez, "Fiscal Volatility Shocks and Economic Activity," Technical Report 11 November 2015.

Fisher, Jonas D.M. and Ryan Peters, "Using Stock Returns to Identify Government Spending Shocks," *The Economic Journal*, 2010, *120* (544), 414–436.

Francis, Neville and Valerie A. Ramey, "Measures of per Capita Hours and Their Implications for the Technology-Hours Debate," *Journal of Money, Credit and Banking*, 09 2009, *41* (6), 1071–1097.

Geweke, John, "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 1999, *18* (1), 1–73.

_ , *Contemporary Bayesian Econometrics and Statistics* Wiley Series in Probability and Statistics, Wiley, 2005.

Guerron-Quintana, Pablo A., "What You Match Does Matter: The Effects of Data on DSGE Estimation," *Journal of Applied Econometrics*, 2010, *25* (5), 774–804.

Gürkaynak, Refet S., Brian Sack, and Eric Swanson, "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models," *American Economic Review*, March 2005, *95* (1), 425–436.

Herbst, Edward P., and Frank Schorfheide, "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, forthcoming.

Iskrev, Nikolay, "Choosing the Variables to Estimate Singular DSGE Models: Comment," Dynare Working Papers 41, CEPREMAP, October 2014.

Koopman, Siem Jan, "Disturbance Smoother for State Space Models," *Biometrika*, 1993, *80* (1), 117–126.

Kuttner, Kenneth N., "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market," *Journal of Monetary Economics*, 2001, *47* (3), 523–544.

Lanne, Markku, Mika Meitz, and Pentti Saikkonen, "Identification and Estimation of Non-Gaussian Structural Vector Autoregressions," CREATES Research Papers 2015-16, School of Eco-nomics and Management, University of Aarhus March 2015.

Leeper, Eric M., Michael Plante, and Nora Traum, "Dynamics of Fiscal financing in the United States," *Journal of Econometrics*, June 2010, *156* (2), 304–321.

Liu, Jun S., "Metropolized Gibbs sampler: An improvement," 1995. unpublished, Department of Statistics, Stanford University.

Lopes, Hedibert F., and Nicholas G. Polson, "Bayesian Instrumental Variables: Priors and Likelihoods," *Econometric Reviews*, June 2014, *33* (1-4), 100–121.

Meng, Xiao-Li, and Wing Hung Wong, "Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration," *Statistica Sinica*, 1996, 831–860.

Mertens, Karel, and Morten O. Ravn, "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review*, June 2013, *103* (4), 1212–1247.

Moon, Hyungsik Roger, and Frank Schorfheide, "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 2012, *80* (2), 755–782.

Ramey, Valerie A., "Identifying Government Spending Shocks: It's All in the Timing," *The Quarterly Journal of Economics*, 2011, *126* (1), 1–50.

Robert, Christian P., and George Casella, *Monte Carlo Statistical Methods* Springer Texts in Statistics, Springer, 2005.

Romer, Christina D., and David H. Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, September 2004, *94* (4), 1055–1084.

_ and _ , "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, June 2010, *100* (3), 763–801.

_ and _ , "Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Changes, 1952–1991," Working Paper 20087, National Bureau of Economic Research, May 2014.

Rossi, Barbara, and Sarah Zubairy, "What Is the Importance of Monetary and Fiscal Shocks in Explaining U.S. Macroeconomic Fluctuations?," *Journal of Money, Credit and Banking*, 2011, *43* (6), 1247–1270.

Rossi, Peter E., Greg M. Allenby, and Rob McCulloch, *Bayesian Statistics and Marketing* Wiley Series in Probability and Statistics, Wiley, 2005.

Rudebusch, Glenn D., "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review*, November 1998, *39* (4), 907–931.

Sims, Christopher A., "Comment on Glenn Rudebusch's "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review*, November 1998, *39* (4), 933–941.

_ , "The State of Macroeconomic Policy Modeling: Where Do We Go from Here?" 2005. accessed 02/12/14.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 2007, *97* (3), 586–606.

Stock, James H. and Mark W. Watson, "Disentangling the Channels of the 2007-2009 Recession," *Brookings Papers on Economic Activity*, 2012, 81–141.

Theil, H. and A. S. Goldberger, "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, 1961, *2* (1), 65–78.

Uhlig, Harald, "What Macroeconomists Should Know about Unit Roots: A Bayesian Perspective," *Econometric Theory*, 1994, *10* (3/4), 645–671.

_ , "What Moves GNP?" 2003, *draft, Humboldt Universit¨at zu Berlin*, 2003.

_ , "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, March 2005, *52* (2), 381–419.

Waggoner, Daniel F. , and Tao Zha, "Confronting Model Misspecification in Macroeconomics," *Journal of Econometrics*, 2012, *171* (2), 167–184. Bayesian Models, Methods and Applications.

Watson, Mark W., "Vector Autoregressions and Cointegration," in R. F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4 of *Handbook of Econometrics*, Elsevier, 1994, chapter 47, 2843–2915.

Wieland, Volker, Tobias Cwik, Gernot J. Müller, Sebastian Schmidt, and Maik Wolters, "A New Comparative Approach to Macroeconomic Modeling and Policy Analysis," *Journal of Economic Behavior & Organization*, 2012, *83* (3), 523–541.

Zha, Tao, "Block Recursion and Structural Vector autoregressions," *Journal of Econometrics*, June 1999, *90* (2), 291–316.

Technical Appendix
– not for publication –

A Narrative VAR and DSGE-VAR

A.1 Narrative shock identification

Here I derive how the observables, Γ and Σ, identify the impulse responses of interest up to an extra $\frac{m_z(m_z-1)}{2}$ restrictions, where m_z is the number of instruments and shocks to identify.

Define

$$\kappa = (\Gamma_1^{-1}\Gamma_2)', \tag{A.1}$$

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

$$\Sigma = \begin{bmatrix} \alpha_{11}\alpha_{11}' + \alpha_{12}\alpha_{12}' & \alpha_{11}\alpha_{11}'\kappa' + \alpha_{12}\alpha_{22}' \\ \kappa\alpha_{11}\alpha_{11}' + \alpha_{22}\alpha_{12}' & \kappa\alpha_{11}\alpha_{11}'\kappa' + \alpha_{22}\alpha_{22}' \end{bmatrix} \tag{A.2}$$

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

$$u_t = A\epsilon_t = \begin{bmatrix} \alpha^{[1]} & \alpha^{[2]} \end{bmatrix} \epsilon_t = \alpha^{[1]}\epsilon_t^{[1]} + \alpha^{[2]}\epsilon_t^{[2]} = \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11}\epsilon_t^{[1]} + \begin{bmatrix} \alpha_{12} \\ \alpha_{22} \end{bmatrix} \epsilon_t^{[2]}$$

Given that $\epsilon^{[1]} \perp\!\!\!\perp \epsilon^{[2]}$, it follows that:

$$\begin{aligned} \mathrm{Var}[u_t|\epsilon_t^{[1]}] &= \alpha^{[2]}(\alpha^{[2]})' = \begin{bmatrix} \alpha_{12}\alpha_{12}' & \alpha_{12}\alpha_{22}' \\ \alpha_{22}\alpha_{12}' & \alpha_{22}\alpha_{22}' \end{bmatrix} \\ \mathrm{Var}[u_t|\epsilon_t^{[2]}] &= \alpha^{[1]}(\alpha^{[1]})' = \begin{bmatrix} \alpha_{11}\alpha_{11}' & \alpha_{11}\alpha_{11}\kappa \\ \kappa\alpha_{11}\alpha_{11}' & \kappa\alpha_{11}\alpha_{11}'\kappa \end{bmatrix} \\ \Sigma = \mathrm{Var}[u_t] &= \mathrm{Var}[u_t|\epsilon_t^{[2]}] + \mathrm{Var}[u_t|\epsilon_t^{[1]}] = \begin{bmatrix} \Sigma_{12}\Sigma_{12}' & \Sigma_{12}\Sigma_{22}' \\ \Sigma_{22}\Sigma_{12}' & \Sigma_{22}\Sigma_{22}' \end{bmatrix} \end{aligned}$$

Note that:

$$u_t^{res} \equiv u_t - \mathbb{E}[u_t|\epsilon_t^{[1]}] \perp \mathbb{E}[u_t|\epsilon_t^{[1]}] = \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11}\epsilon_t^{[1]}$$

Any vector in the nullspace of $\begin{bmatrix} I_{m_z} & \kappa' \end{bmatrix}$ satisfies the orthogonality condition.

Note that $\left\{ \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix}, \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \right\}$ is an orthogonal basis for $\mathbb{R}^m$.

Define

$$Z \equiv \begin{bmatrix} Z^{[1]} & Z^{[2]} \end{bmatrix} \equiv \begin{bmatrix} I_{m_z} & \kappa' \\ \kappa & -I_{m-m_z} \end{bmatrix} \tag{A.3}$$

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 $\epsilon_t^{[1]}$.

$$\begin{aligned} (Z^{[2]})'v_t &= (Z^{[2]})'A\epsilon_t = (Z^{[2]})' \begin{bmatrix} \alpha^{[1]} & \alpha^{[2]} \end{bmatrix} \epsilon_t \\ &= (Z^{[2]})' \begin{bmatrix} Z^{[1]} \| \tilde{Z} \| \alpha_{11} & \alpha^{[2]} \end{bmatrix} \epsilon_t = \begin{bmatrix} 0 & (Z^{[2]})'\alpha^{[2]} \end{bmatrix} \epsilon_t \end{aligned}$$

$$= 0 \times \epsilon_t^{[1]} + (Z^{[2]})'\alpha^{[2]}\epsilon_t^{[2]} \perp\!\!\!\perp \epsilon^{[1]}$$

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:

$$\begin{aligned}
\mathbb{E}[v_t|\epsilon_t^{[2]}] &= \text{Cov}[v_t, (Z^{[2]})'v_t]\,\text{Var}[(Z^{[2]})'v_t]^{-1}(Z^{[2]})'v_t, \\
v_t - \mathbb{E}[v_t|\epsilon_t^{[2]}] &= (I - \text{Cov}[v_t, (Z^{[2]})'v_t]\,\text{Var}[(Z^{[2]})'v_t]^{-1}(Z^{[2]})')v_t, \\
\text{Cov}[v_t, (Z^{[2]})'v_t] &= \Sigma Z^{[2]} = \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \\
\text{Var}[v_t|\epsilon_t^{[2]}] &= \mathbb{E}[(I - \text{Cov}[v_t, (Z^{[2]})'v_t]\,\text{Var}[(Z^{[2]})'v_t]^{-1}(Z^{[2]})')v_t v_t'] \\
&= \mathbb{E}[v_t v_t'] - \text{Cov}[v_t, (Z^{[2]})'v_t]\,\text{Var}[(Z^{[2]})'v_t]^{-1}\mathbb{E}[(Z^{[2]})')v_t v_t'] \\
&= \Sigma - \text{Cov}[v_t, (Z^{[2]})'v_t]\,\text{Var}[(Z^{[2]})'v_t]^{-1}\,\text{Cov}[v_t, (Z^{[2]})'v_t] \\
&= \Sigma - \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \left(\begin{bmatrix} \kappa & -I_{m-m_z} \end{bmatrix} \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \right)^{-1} \begin{bmatrix} \kappa & -I_{m-m_z} \end{bmatrix} \Sigma \\
&= \begin{bmatrix} \alpha_{11}\alpha_{11}' & \alpha_{11}\alpha_{11}\kappa \\ \kappa\alpha_{11}\alpha_{11}' & \kappa\alpha_{11}\alpha_{11}'\kappa \end{bmatrix} \qquad \text{(A.4)}
\end{aligned}$$

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

$$\begin{aligned}
\text{Var}[v_t|\epsilon_t^{[1]}] &= \Sigma - \text{Var}[v_t|\epsilon_t^{[2]}] \\
&= \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \left(\begin{bmatrix} \kappa & -I_{m-m_z} \end{bmatrix} \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \right)^{-1} \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix}' \Sigma \qquad \text{(A.5)}
\end{aligned}$$

In general, α_{11} itself is unidentified: 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 $\alpha_{11}\alpha_{11}'$. Given α_{11}, the impact response to a unit shock is given by:

$$\begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11}$$

A.2 Narrative policy rule identification

To show that the lower Cholesky 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 (3.8). Recall the definition of forecast errors v_t in terms of structural shocks ϵ_t:

$$v_t = A\epsilon_t \equiv \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}^{-1} \qquad v_t = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \qquad \text{(A.6)}$$

Note that:

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

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

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:

$$S_1 \equiv (I - \alpha_{12}\alpha_{22}^{-1}\alpha_{21}\alpha_{11}^{-1})\alpha_{11} \quad S_2 \equiv (I - \alpha_{21}\alpha_{11}^{-1}\alpha_{12}\alpha_{22}^{-1})\alpha_{22} \tag{A.7}$$

so that

$$(\alpha_{11} - \alpha_{12}\alpha_{22}^{-1}\alpha_{21})^{-1} = S_1^{-1} \quad (\alpha_{22} - \alpha_{21}\alpha_{11}^{-1}\alpha_{12})^{-1} = S_2^{-1}$$

Using these equalities gives the first equality in what follows, whereas the second equality is straightforward algebra:

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}^{-1} v_t = \begin{bmatrix} S_1^{-1} & -S_1^{-1}\alpha_{11}^{-1}\alpha_{12} \\ -S_2^{-1}\alpha_{22}^{-1}\alpha_{21} & S_2^{-1} \end{bmatrix} v_t$$
$$= \begin{bmatrix} S_1^{-1} & \mathbf{0} \\ \mathbf{0} & S_2^{-1} \end{bmatrix} \begin{bmatrix} I & -\alpha_{12}\alpha_{22}^{-1} \\ -\alpha_{21}\alpha_{11}^{-1} & I \end{bmatrix} v_t = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

and equivalently:

$$\begin{bmatrix} I & -\eta \\ -\kappa & I \end{bmatrix} v_t = \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & S_2 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{A.8}$$

defining $\eta \equiv \alpha_{12}\alpha_{22}^{-1}$ and $\kappa \equiv \alpha_{21}\alpha_{11}^{-1}$. Equation (3.8) follows immediately.

Lemma 4 (Mertens and Ravn (2013)). *Let $\Sigma = AA'$ and $\Gamma = \begin{bmatrix} G & \mathbf{0} \end{bmatrix} A$, 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 $S_1 S_1'$ with S_1 defined in (A.7).*

Proof. Since A is of full rank, it is invertible, and (A.8) holds for any such A. Given η, κ, (A.8) implies (3.9), which I reproduce here for convenience:

$$\alpha^{[1]} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \begin{bmatrix} (I - \eta\kappa)^{-1} \\ (I - \kappa\eta)^{-1}\kappa \end{bmatrix} \text{chol}(S_1 S_1'). \tag{3.9}$$

If Σ and Γ pin down η, κ uniquely, $\alpha^{[1]}$ is uniquely identified except for a factorization of $S_1 S_1'$.

To show that Σ and Γ pin down η, κ uniquely, consider κ first. Since $\Gamma = \begin{bmatrix} G & \mathbf{0} \end{bmatrix} A$ and G is an $m_z \times m_z$ invertible matrix, it follows that Assumption 1 holds. It then follows from (3.3) that $\kappa = \alpha_{21}\alpha_{11}^{-1} = \Gamma^2\Gamma_1^{-1}$.

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

$$\alpha_{22}\alpha_{22}' = \Sigma_{22} - \kappa\alpha_{11}\alpha_{11}'\kappa' = \Sigma_{22} - \kappa(\Sigma_{11} - \alpha_{12}\alpha_{12}')\kappa',$$

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

$$\alpha_{12}\alpha_{12}' = (\Sigma_{12}' - \kappa\Sigma_{11})'(ZZ')^{-1}(\Sigma_{12}' - \kappa\Sigma_{11})$$

with

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

The coefficient matrix of interest, η, is then defined as:

$$\begin{aligned}\eta &\equiv \alpha_{12}\alpha_{22}^{-1} = \alpha_{12}\alpha_{22}'(\alpha_{22}\alpha_{22}')^{-1} = (\Sigma_{12}' - \kappa\alpha_{11}\alpha_{11}')'(\alpha_{22}\alpha_{22}')^{-1} \\ &= (\Sigma_{12}' - \kappa\Sigma_{11}' + \kappa\alpha_{12}\alpha_{12}')'(\alpha_{22}\alpha_{22}')^{-1}.\end{aligned}$$

Thus, η and κ are uniquely identified given Σ, Γ. □

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 $\Gamma = [G, \boldsymbol{0}]A^*$. Then $\alpha^{[1]}$ defined in (3.9) satisfies $\alpha^{[1]} = A^*[I_{m_z}, \boldsymbol{0}_{(m-m_z)\times(m-m_z)}]'$ 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 4, $\alpha^{[1]}$ 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.7) for arbitrary full rank A, is unique up to a normalization. It then follows that $\alpha^{[1]}$ is identified uniquely and, hence, equal to $A^*[I_{m_z}, \mathbf{0}_{(m-m_z)\times(m-m_z)}]'$.

To proceed, stack the m_p policy rules:

$$\begin{aligned} Y_t^p &= \sum_{i=m_p+1}^{m} \begin{bmatrix} \psi_{1,i} & 0 & \ldots & 0 \\ 0 & \psi_{2,i} & \ldots & 0 \\ \vdots & \ldots & \ddots & \vdots \\ 0 & 0 & \ldots & \psi_{m_p,i} \end{bmatrix} y_{i,t} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n_p} \end{bmatrix} X_{t-1} + \begin{bmatrix} \sigma_{11} & 0 & \ldots & 0 \\ \sigma_{21} & \sigma_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_p,1} & \sigma_{n_p,2} & \ldots & \sigma_{n_p,n_p} \end{bmatrix} \epsilon_t^p \\ &\equiv \sum_{i=m_p+1}^{m} D_i y_{i,t} + \Lambda X_{t-1} + \begin{bmatrix} D_0 & \mathbf{0} \end{bmatrix} \epsilon_t, \\ &= \left(\begin{bmatrix} D_0 & \mathbf{0} \end{bmatrix} \epsilon_t + \sum_{i=m_p+1}^{m} D_i \mathbf{1} A_i' \right) \epsilon_t + \left(\sum_{i=m_p+1}^{m} D_i \mathbf{1} B_i' X_{t-1} + \Lambda \right) X_{t-1}, \end{aligned}$$

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

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

$$\begin{bmatrix} [D_0, \mathbf{0}] + \sum_{i=m_p}^{m} D_i \mathbf{1} (A_i^*)' \\ (A_{m_p+1}^*)' \\ \vdots \\ (A_m^*)' \end{bmatrix},$$

where D_0 is a full-rank lower diagonal matrix and the D_j matrices are $m_p \times m_p$ matrices.

To find $(A^*)^{-1}$, proceed by Gauss-Jordan elimination to rewrite the system $A^*X = I_m$, with solution $X = (A^*)^{-1}$, as $[A^*|I_m]$. Define E as a conformable matrix such that $[A^*|I_m] \overset{E}{\to} [B|C] = [EA^*|EI_m]$. Then:

$$[(A^*)|I_m] = \left[\begin{array}{c|ccccc} [D_0 \ \ \mathbf{0}] + \sum_{i=m_p+1}^{m} D_i \mathbf{1}(A_i^*)' & I_{m_p} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ (A^*_{m_p+1})' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (A^*_m)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right]$$

$$\overset{E_1}{\to} \left[\begin{array}{c|ccccc} [D_0 \ \ \mathbf{0}] + \sum_{i=m_p+2}^{m} D_i \mathbf{1}(A_i^*)' & I_{m_p} & -D_{m_p+1}\mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ (A^*_{m_p+1})' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (A^*_{m_p+2})' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (A^*_m)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right]$$

$$\overset{E_2}{\to} \left[\begin{array}{c|ccccc} [D_0 \ \ \mathbf{0}] + \sum_{i=m_p+3}^{m} D_i \mathbf{1}(A_i^*)' & I_{m_p} & -D_{m_p+1}\mathbf{1} & -D_{m_p+2}\mathbf{1} & \dots & \mathbf{0} \\ (A^*_{m_p+1})' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (A^*_{m_p+2})' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (A_m *)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right] \overset{E_3}{\to} \dots$$

$$\overset{E_{m-m_p}}{\to} \left[\begin{array}{c|ccccc} [D_0 \ \ \mathbf{0}] & I_{m_p} & -D_{m_p+1}\mathbf{1} & -D_{m_p+2}\mathbf{1} & \dots & -D_m\mathbf{1} \\ (A^*_{m_p+1})' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (A^*_{m_p+2})' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (A_m *)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right]$$

$$\overset{E_D}{\to} \left[\begin{array}{c|ccccc} [I_{m_p} \ \ \mathbf{0}] & D_0^{-1} & -D_0^{-1}D_{m_p+1}\mathbf{1} & -D_0^{-1}D_{m_p+2}\mathbf{1} & \dots & -D_0^{-1}D_m\mathbf{1} \\ (A^*_{m_p+1})' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (A^*_{m_p+2})' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (A_m *)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right]$$

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

Now consider cases (a) and (b):

(a) $m_z = m_p$. From (A.7), 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 (lower) diagonal matrix because D_0 is (lower) diagonal.

(b) $m_z = m_p + 1$, $\psi_{p,m_p+1} = 0, p = 1, \dots, m_p$. The second condition implies that $D_{m_p+1} = \mathbf{0}_{m_p \times m_p}$. It follows that S_1 defined in (A.7) is given by:

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

Thus, S_1 is lower triangular.

In both cases, S_1 is lower triangular. Since the lower Cholesky decomposition is unique, a Cholesky 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 4. □

A.3 Priors and posteriors

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

$$\begin{aligned} L &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} u_t' V^{-1} u_t\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \operatorname{tr}(u_t' V^{-1} u_t)\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{tr}(V^{-1}\sum_{t=1}^{T} u_t u_t')\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{tr}(V^{-1}U'U)\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{vec}(U)'(V^{-1}\otimes I_T)\operatorname{vec}(U)\right), \end{aligned} \tag{A.9}$$

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

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

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

The likelihood can then also be written as:

$$\begin{aligned} L &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}(Y_{SUR} - X_{SUR}\beta)'(V^{-1}\otimes I_T)(Y_{SUR} - X_{SUR}\beta)\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}(\tilde{Y}_{SUR} - \tilde{X}_{SUR}\beta)'(\tilde{Y}_{SUR} - \tilde{X}_{SUR}\beta)\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}(\tilde{Y}_{SUR} - \tilde{X}_{SUR}\beta)'(\tilde{Y}_{SUR} - \tilde{X}_{SUR}\beta)\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}(\tilde{X}_{SUR}(\beta - \tilde{\beta}_{SUR}))'(\tilde{X}_{SUR}(\beta - \tilde{\beta}_{SUR}))\right) \\ &= (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}(\beta - \tilde{\beta}_{SUR})'(\tilde{X}_{SUR}'\tilde{X}_{SUR})(\beta - \tilde{\beta}_{SUR})\right), \end{aligned} \tag{A.10}$$

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

Note that expression (A.10) for the likelihood is proportional to a conditional Wishart distribution for β: $\beta|V^{-1} \sim \mathcal{N}(\tilde{\beta}_{SUR}, (\tilde{X}_{SUR}'\tilde{X}_{SUR})^{-1}) \equiv \mathcal{N}(\tilde{\beta}_{SUR}, (X_{SUR}'(V^{-1}\otimes I)X_{SUR})^{-1})$. Alternatively, expression (A.9) for the likelihood is proportional to a conditional Wishart distribution for V^{-1}: $V^{-1}|\beta \sim \mathcal{W}_{m_a}((U(\beta)'U(\beta))^{-1}, T + m_a + 1)$. Premultiplying with a Jeffrey'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:

$$\begin{aligned} \pi(V^{-1})\times &= |V^{-1}|^{-\frac{m_a+1}{2}} \times (2\pi)^{-mT/2}|V|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{tr}(V^{-1}U'U)\right) \\ &= (2\pi)^{-mT/2}|V^{-1}|^{(T-m_a-1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}(V^{-1}U'U)\right), \end{aligned} \tag{A.11}$$

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

$$\begin{aligned} SSR(\beta) \equiv U(\beta)'U(\beta) &= [Y - X_y B_y(\beta), Z - X_z B_z(\beta)]'[Y - X_y B_y(\beta), Z - X_z B_z(\beta)] \\ &= \sum_{t=1}^{T} [y_t - x_{y,t} B_y(\beta), z_t - x_{z,t} B_z(\beta)]'[y_t - x_{y,t} B_y(\beta), z_t - x_{z,t} B_z(\beta)]. \end{aligned}$$

A.4 Dummy variables prior

Note that the dummy variables prior is no longer conjugate. Hence, my prior can be generated from two different distributions: The coefficients are generated from a $\mathcal{N}(\beta_0, \bar{V}_0^{-1})$ distribution, whereas the observations that generate the prior for the covariance matrix are generated from a $\mathcal{N}(0, V^{-1})$ distribution.

Specify:

$$\beta \sim \mathcal{N}(\bar{\beta}_0, \bar{N}_0^{-1}), \qquad \bar{N}_0 \equiv X'_{SUR,0}(\bar{V}_0^{-1} \otimes I) X_{SUR,0}$$

Note that this is not equal to

$$\beta|V^{-1} \sim \mathcal{N}(\bar{\beta}_0, \bar{N}_0(V^{-1})), \qquad \bar{N}_0(V^{-1}) \equiv X'_{SUR,0}(V^{-1} \otimes I) X_{SUR,0},$$

unless V^{-1} is known and equal to $\bar{V}_0$.

The prior for V^{-1} is Wishart independent of β.

$$V^{-1} \sim \mathcal{W}_{m+m_z}(\bar{V}_0 T_0, T_0)$$

Note that because the prior for β is independent of V^{-1}, the prior is conditionally conjugate with the likelihood function. Otherwise, the presence of $|N_0(V^{-1})|$ terms would undo the conjugacy.

The prior is therefore:

$$\begin{aligned} \pi(\beta, V^{-1}|\theta) &= (2\pi)^{-n/2}|\bar{N}_0(\theta)|^{+1/2} e^{-\frac{1}{2}(\beta - \bar{\beta}_0(\theta))'\bar{N}_0(\theta)(\beta - \bar{\beta}_0(\theta))} \\ &\quad \times 2^{-T_0(m+m_z)/2}|\bar{V}_0(\theta)T_0|^{-T_0/2}\Gamma_m(T_0/2)^{-1}|V^{-1}|^{(T_0-m-m_z-1)/2} e^{-\frac{1}{2}\operatorname{tr}(V^{-1}\bar{V}_0(\theta)T_0)} \\ &= (2\pi)^{-n/2}|\bar{N}_0(\theta)|^{+1/2} e^{-\frac{1}{2}(\beta - \bar{\beta}_0(\theta))'\bar{N}_0(\theta)(\beta - \bar{\beta}_0(\theta))} \\ &\quad \times 2^{-T_0(m+m_z)/2}|S_0(\theta)|^{-T_0/2}\Gamma_m(T_0/2)^{-1}|V^{-1}|^{(T_0-m-m_z-1)/2} e^{-\frac{1}{2}\operatorname{tr}(V^{-1}S_0(\theta))} \end{aligned}$$

The joint density is given by:

$$\begin{aligned} p(Y, Z, \beta, V^{-1}, \theta) &= p(Y, Z|\beta, V^{-1})p(\beta, V^{-1}|\theta)p(\theta), \\ &= p(Y, Z|\beta, V^{-1})p(\beta|\theta)p(V^{-1}|\theta)p(\theta), \end{aligned} \tag{A.12a}$$

$$p(Y, Z|\beta, V^{-1}) = (2\pi)^{-T/2}|V^{-1}|^{T/2} e^{-\frac{1}{2}(\operatorname{vec}([Y,Z]-X_{SUR}\beta)'(V^{-1}\otimes I_T)(\operatorname{vec}([Y,Z]-X_{SUR}\beta)} \tag{A.12b}$$

$$p(\beta|\theta) = (2\pi)^{-n_\beta/2}\left|\lambda_B X'_{0,SUR}(\theta)(\bar{V}_0(\theta)^{-1} \otimes I_{m(mp+k)})X_{0,SUR}\right|^{1/2}$$

$$e^{-\frac{\lambda_B}{2}(X_{0,SUR}(\bar{\beta}_0(\theta)-\beta))'(\bar{V}_0^{-1}\otimes I_{m(mp+k)})(X_{0,SUR}(\bar{\beta}_0(\theta)-\beta))},$$

$$\text{where } \lambda_B \equiv \frac{T_0^B}{m(mp+k)}$$

$$= (2\pi)^{-n_\beta/2}|\bar{N}_0(\theta)|^{1/2}e^{-\frac{1}{2}(\bar{\beta}_0(\theta)-\beta)'\bar{N}_0(\theta)(\bar{\beta}_0(\theta)-\beta)}, \tag{A.12c}$$

$$p(V^{-1}|\theta) = \frac{e^{-\frac{1}{2}\operatorname{tr}(V^{-1}T_0^V\bar{V}_0(\theta))}}{2^{T_0^V(m+m_z)/2}\Gamma_{m+m_z}\left(\frac{T_0^V}{2}\right)}\frac{|T_0^V\bar{V}_0(\theta)|^{T_0^V/2}}{|V|^{(T_0^V-m-m_z-1)/2}} \tag{A.12d}$$

$$p(\theta) = \mathbf{1}\{\text{DSGE model has a unique \& stable solution}|\theta\} \times \prod_{n=1}^{n_\theta} p_n(\theta^{(n)}), \tag{A.12e}$$

where $\theta^{(n)}$ denotes the nth component of the vector θ and $p_n(\theta^{(n)})$ is a univariate density.

A.5 Likelihood computation

I compute the marginal data density by applying the Chib (1995) method to the inner integral over the SUR-VAR parameters and then applying the Geweke (1999) estimator to integrate over the DSGE model hyperparameters.

The basic insight from Chib (1995) is that:

$$\pi(y,Z) = \frac{p(y,Z)\pi(V^{-1},B|\theta)\pi(\theta)}{\pi(\theta,V^{-1},B|y,Z)} = \frac{p(y,Z)\pi(V^{-1},B|\theta)\pi(\theta)}{\pi(\theta|y,Z)\pi(V^{-1}|y,Z,\theta)\pi(B|y,Z,\theta,V^{-1})}, \tag{A.13}$$

for any θ. For numerical purposes, however, it is advisable to evaluate (A.13) at a high density point. In what follows I denote this point by $(\theta_*, B_*, V_*^{-1})$.

Geweke (1999) shows that to find the integrating constant of a Kernel $k(\psi)$, we may use that $p(\tilde{y})$ is the integrating constant of the posterior kernel $k(\psi) = p(\psi|\tilde{y})p(\tilde{y})$. Let $g(\psi) \equiv \frac{f(\psi)}{k(\psi)}$. Then:

$$\mathbb{E}[g(\psi)] = \int_\Psi \frac{f(\psi)}{k(\psi)}p(\psi|\tilde{y})d\psi = p(\tilde{y})^{-1}\int_\Psi \frac{f(\psi)}{k(\psi)}k(\psi)d\psi = p(\tilde{y})^{-1}\int_\Psi f(\psi)d\psi = p(\tilde{y})^{-1}$$

for any density $f(\psi)$. Geweke (1999) proposes to use a truncated normal density function with the posterior mean and covariance of ψ. Denote this truncated density by $f_\alpha(\psi)$ and its estimate based on the sample posterior distribution with sample size M by $f_{\alpha,M}(\psi)$. Then:

$$p(\tilde{y})^{-1} = \mathbb{E}[g_\alpha(\psi)] \approx \mathbb{E}[g_{\alpha,M}(\psi)] \approx M^{-1}\sum_{m=1}^{M}\frac{f_{\alpha,M}(\psi_m)}{k(\psi_m)},$$

where ψ_m are draws from the posterior.

Here, $\psi = (\theta, B, V^{-1})$ – or strictly $(\theta, B, \operatorname{vech}(V^{-1}))$. This vector is high dimensional, especially because of the presence of B. It would therefore be helpful to reduce the dimensionality of the parameter vector.

$$k(\theta,B,V^{-1}) = p(y,z|B,V^{-1})p(B,V^{-1}|\theta)\pi(\theta)$$

$$\Rightarrow k(\theta) \equiv \int\int k(\theta,B,V^{-1})dV^{-1}dB = \pi(\theta)\int\int p(y,z|B,V^{-1})p(B,V^{-1}|\theta)dV^{-1}dB$$

$$\equiv \pi(\theta)\int\int k(B,V^{-1}|y,z,\theta)dV^{-1}dB$$

$$= \pi(\theta)p(y,z,\theta)\int\int p(B,V^{-1}|y,z,\theta)dV^{-1}dB$$
$$\Leftrightarrow k(\theta) = p(y,z,\theta)\pi(\theta)$$

Now proceed with this reduced parameter vector as before

$$\mathbb{E}[g(\theta)] = \int_\Theta \frac{f(\theta)}{\pi(\theta)p(y,z,\theta)}p(\theta|y,z)d\theta = \int_\Theta \frac{f(\theta)}{\pi(\theta)p(y,z,\theta)}p(\theta|y,z)d\theta$$
$$= p(y,z)^{-1}\int_\Theta \frac{f(\theta)}{\pi(\theta)p(y,z,\theta)}\pi(\theta)p(y,z,\theta)d\theta$$
$$= p(y,z)^{-1}\int_\Theta f(\theta)d\theta = p(y,z)^{-1}$$

In practice, I approximate $p(y,z|\theta)$ with the Chib (1995) estimator:

$$\hat{\mathbb{E}}[\hat{g}(\theta)] = \frac{1}{M}\sum_{m=1}^{M}\frac{f_\alpha(\theta_{(m)})}{\pi(\theta_{(m)})\hat{p}(y,z,\theta_{(m)})} \approx \hat{p}(y,z)^{-1}$$

where $\hat{p}(y,z,\theta_{(m)})$ is the Chib estimator of the (conditional) marginal likelihood. The approximation relies on $\int_\Theta f(\theta)\frac{p(y,z,\theta)}{\hat{p}(y,z,\theta)}d\theta$ being small. In the case without instruments and with a fully conjugate prior, I verify this numerically by comparing the estimated marginal data density with its analytical counterpart. For the SUR case, I verify that with a modest number of posterior draws the numerical error lies within ± 0.1 of the truth computed by a very large number of draws.

A.5.1 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:

$$\tilde{Y}_t = \sum_{l=1}^{k} A_l \tilde{Y}_{t-l} + \epsilon_t, \quad \tilde{Y}_t \equiv Y_t - \tilde{D}_y t - \tilde{\mu}_y. \tag{A.14}$$

$$\Leftrightarrow \quad Y_t = \left(\sum_{l=1}^{k} A_l Y_{t-l}\right) + \left(I - \sum_{l=1}^{k} A_l\right)\tilde{D}_y t + \left(\tilde{\mu}_y - \sum_{l=1}^{k} A_l(\tilde{\mu}_y - \tilde{D}_y \times l)\right) + \epsilon_t$$

$$\equiv \left(\sum_{l=1}^{k} A_l Y_{t-l}\right) + D_y t + \mu_y + \epsilon_t \tag{A.15}$$

Equation (A.15) 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 $\tilde{\mu}_y = \left(I - \sum_{l=1}^{k} A_l\right)^{-1}\left(\mu_y - \sum_{l=1}^{k} A_l \tilde{D}_y \times l\right)$.

To compute variances and covariances as required to implement the dummy variable prior, I first compute the population moments of $\tilde{Y}_t$ in (A.14), using that $\mathbb{E}[\tilde{Y}_t] = 0$. 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.[44] 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.

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

A.5.2 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 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):

$$z_{i,t} = c_i \epsilon_{i,t} + \sum_{j \neq i} c_j \epsilon_{j,t} + u_{i,t}, \quad u_{i,t} \overset{iid}{\sim} \mathcal{N}(0,1).. \tag{A.16}$$

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 Normal prior for c_j centered at zero. In particular, $c_j \sim \mathcal{N}(0, \frac{1}{2^2})$. For c_i, I consider both $c_i \sim \mathcal{N}(1, \frac{1}{2^2})$ and $c_i = 1$ after scaling the instrument to be measured in units of $\epsilon_{i,t}$.

A.6 DSGE model equations

A.6.1 Households

The law of motion for capital:

$$\hat{k}_t^p = (1 - \frac{\bar{x}}{\bar{k}^p})\hat{k}_{t-1}^p + \frac{\bar{x}}{\bar{k}^p}(\hat{x}_t + \hat{q}_{t+s}) \tag{A.17}$$

Household wage setting:

$$\begin{aligned}
\hat{w}_t = & \frac{\hat{w}_{t-1}}{1+\bar{\beta}\gamma} + \frac{\bar{\beta}\gamma \mathbb{E}_t[\hat{w}_{t+1}]}{1+\bar{\beta}\gamma} \\
& + \frac{(1-\bar{\beta}\zeta_w\gamma)(1-\zeta_w)}{(1+\bar{\beta}\gamma)\zeta_w}\bar{A}_w\Big(\frac{\hat{c}_t - (h/\gamma)\hat{c}_{t-1}}{1-h/\gamma} + \nu\hat{n}_t - \hat{w}_t + \frac{d\tau_t^n}{1-\bar{\tau}^n} + \frac{d\tau_t^c}{1+\bar{\tau}^c}\Big) \\
& - \frac{1+\bar{\beta}\mu\iota_w}{1+\bar{\beta}\gamma}\hat{\pi}_t + \frac{\iota_w}{1+\bar{\beta}\gamma}\hat{\pi}_{t-1}\frac{\bar{\beta}\gamma}{1+\bar{\beta}\gamma}\mathbb{E}_t[\hat{\pi}_{t+1}] + \frac{\hat{\epsilon}_t^{\lambda,w}}{1+\bar{\beta}\gamma}
\end{aligned} \tag{A.18}$$

Household consumption Euler equation:

$$\begin{aligned}
& \mathbb{E}_t[\hat{\xi}_{t+1} - \hat{\xi}_t] + \mathbb{E}_t[d\tau_{t+1}^c - d\tau_t^c] = \\
& = \frac{1}{1-h/\gamma}\mathbb{E}_t\big((\sigma-1)\frac{1}{1+\bar{\lambda}_w}\frac{1-\bar{\tau}^n}{1+\tau^c}\frac{\bar{w}\bar{n}}{\bar{c}}[\hat{n}_{t+1} - \hat{n}_t] - \sigma\Big[\hat{c}_{t+1} - \Big(1+\frac{h}{\gamma}\Big)c_t + \frac{h}{\gamma}\hat{c}_{t+1}\Big]\Big),
\end{aligned} \tag{A.19}$$

Other FOC (before rescaling of $\hat{q}_t^b$):

$$\mathbb{E}_t[\hat{\xi}_{t+1} - \hat{\xi}_t] = -\hat{q}_t^b - \hat{R}_t + \mathbb{E}_t[\hat{\pi}_{t+1}], \tag{A.20}$$

$$\begin{aligned}
\hat{Q}_t = & -\hat{q}_t^b - (\hat{R}_t - \mathbb{E}_t[\pi_{t+1}]) + \frac{1}{\bar{r}^k(1-\tau^k) + \delta\tau^k + 1 - \delta} \times \\
& \times \Big[(\bar{r}^k(1-\tau^k) + \delta\tau^k)\hat{q}_t^k - (\bar{r}^k - \delta)d\tau_{t+1}^k +
\end{aligned} \tag{A.21}$$

$$+ \bar{r}^k(1-\tau^k)\mathbb{E}_t(\hat{r}_{t+1}^k) + (1-\delta)\mathbb{E}_t(\hat{Q}_{t+1})\Big], \tag{A.22}$$

$$\hat{x}_t = \frac{1}{1+\bar{\beta}\gamma}\left[\hat{x}_{t-1} + \bar{\beta}\gamma\mathbb{E}_t(\hat{x}_{t+1}) + \frac{1}{\gamma^2 S''(\gamma)}[\hat{Q}_t + \hat{q}_t^x]\right], \tag{A.23}$$

$$\hat{u}_t = \frac{a'(1)}{a''(1)}\hat{r}_t^k \equiv \frac{1-\psi_u}{\psi_u}\hat{r}_t^k. \tag{A.24}$$

A.6.2 Production side and price setting

The linearized aggregate production function is:

$$\hat{y}_t = \frac{\bar{y}+\Phi}{\bar{y}}\left(\hat{\epsilon}_t^a + \zeta\hat{k}_{t-1}^g + \alpha(1-\zeta)\hat{k}_t + (1-\alpha)(1-\zeta)\hat{n}_t\right), \tag{A.25}$$

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

The capital-labor ratio:

$$\hat{k}_t = \hat{n}_t + \hat{w}_t - \hat{r}_t^k. \tag{A.26}$$

Price setting:

$$\hat{\pi}_t = \frac{\iota_p}{1+\iota_p\bar{\beta}\gamma}\hat{\pi}_{t-1} + \frac{1-\zeta_p\bar{\beta}\gamma}{1+\iota_p\bar{\beta}\gamma}\frac{1-\zeta_p}{\zeta_p}\bar{A}_p(\widehat{mc}_t + \hat{\epsilon}_t^{\lambda,p}) + \frac{\bar{\beta}\gamma}{1+\iota_p\bar{\beta}\gamma}\mathbb{E}_t\hat{\pi}_{t+1}. \tag{A.27}$$

Marginal costs with a cost channel:

$$\widehat{mc}_t = \alpha\hat{r}_t^k + (1-\alpha)(\hat{w}_t + \hat{R}_t). \tag{A.28}$$

A.6.3 Market clearing

Goods market clearing requires:

$$\hat{y}_t = \frac{\bar{c}}{\bar{y}}\hat{c}_t + \frac{\bar{x}}{\bar{y}}\hat{x}_t + \frac{\bar{x}^g}{\bar{y}}\hat{x}_t^g + \hat{g}_t + \frac{\bar{r}^k\bar{k}}{\bar{y}}\hat{u}_t. \tag{A.29}$$

A.6.4 Observation equations

For the estimation under full information, I need to specify observation equations. The observation equations are given by (3.2) as well as the following seven observation equations from Smets and Wouters (2007) and three additional equations (A.31) on fiscal variables:

$$\Delta \ln g_t^{obs} = g_t - g_{t+1} + (\gamma_g - 1), \tag{A.30a}$$

$$\Delta \ln x_t^{obs} = x_t - x_{t+1} + (\gamma_x - 1), \tag{A.30b}$$

$$\Delta \ln w_t^{obs} = w_t - w_{t+1} + (\gamma_w - 1), \tag{A.30c}$$

$$\Delta \ln c_t^{obs} = c_t - c_{t+1} + (\gamma - 1), \tag{A.30d}$$

$$\hat{\pi}_t^{obs} = \hat{\pi}_t + \bar{\pi}, \tag{A.30e}$$

$$\hat{n}_t^{obs} = \hat{n}_t + \bar{n}, \tag{A.30f}$$

$$\hat{R}_t^{obs} = \hat{R}_t + (\beta^{-1} - 1), \tag{A.30g}$$

By allowing for different trends in the nonstationary 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:

$$b_t^{obs} = \frac{\bar{b}}{\bar{y}}(\hat{b} - \hat{y}) + \bar{b}^{obs} \tag{A.31a}$$

$$rev_t^{n,obs} = \bar{\tau}^n \frac{\bar{w}\bar{n}}{\bar{c}} \frac{\bar{c}}{\bar{y}} \left(\frac{d\tau_t^n}{\bar{\tau}^n} + \hat{w}_t + \hat{n}_t - \hat{y}_t \right) + r\bar{e}v^{n,obs} \tag{A.31b}$$

$$rev_t^{k,obs} = \bar{\tau}^k \frac{\bar{k}}{\bar{y}}(\bar{r}^k - \delta) \left(\frac{d\tau^k}{\bar{\tau}^k} + \frac{\bar{r}^k}{\bar{r}^k - \delta}\hat{r}_t^k + \hat{k}_{t-1}^p - \hat{y}_t \right) + r\bar{e}v^{k,obs} \tag{A.31c}$$

A.7 Inferring the response of additional variables

Zha (1999) shows that such a block-recursive system can easily be implemented by including current values of the original "core" VAR variables as (exogenous) regressors to a separate periphery-VAR, for which standard inference applies:

$$Y_{p,t} = B_p^p Y_{p,t-1} + B_{c,1}^p Y_{c,t-1} + B_{c,0}^p Y_{c,t} + A_p^p \epsilon_t^p. \tag{A.32}$$

Conditional on current core variables, the additional "peripheral" variables in (A.32) 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):

$$\Sigma_p^{-1} \equiv (A_p^p (A_p^p)')^{-1} \sim \mathcal{W}_n(((\nu_0 + T)S_{p,T})^{-1}, (\nu_0 + T)), \tag{A.33a}$$

$$N_{p,T} = N_{p,0} + X_p' X_p,$$

$$S_{p,T} = \frac{1}{\nu_T}(\hat{B}_p - \bar{B}_{p,0})' N_{p,0} N_{p,T}^{-1} X_p' X_p (\hat{B}_p - \bar{B}_{p,0}) + \frac{\nu_0}{\nu_T} S_0 + \frac{\nu_T - \nu_0}{\nu_T}\hat{\Sigma}_p$$

$$B_p | \Sigma_p \sim \mathcal{N}(B_{p,T}, \Sigma_p \otimes N_{p,T}^{-1}). \tag{A.33b}$$

Here, $\hat{B}_p = (X_p' X_p)^{-1} X_p' Y_p$, $\bar{B}_{p,T} = N_{p,T}^{-1}(N_{p,0}\bar{B}_{p,0} + X_p' X_p \hat{B}_p)$, $\hat{\Sigma}_p = T^{-1} Y_p' (I - X_p (X_p' X_p)^{-1} X_p') Y_p$. 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 $B_{c,0}^p \tilde{A} q_{IV}$ 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 $\left[I_{m+m_p} \mathbf{0}_{(m+m_p)\times(p-1)(m+m_p)} \right] B^h \begin{bmatrix} \tilde{A} q_{IV} \\ \mathbf{0}_{(p-1)(m+m_p)\times m} \end{bmatrix}$. [45]

A.8 Results on the marginal data density in T_0^V

A.8.1 Analytic results

Del Negro et al. (2007) show that, in an AR(1) model with known variance, the marginal likelihood is strictly increasing, decreasing, or has an interior maximum in $T_0^V = T_0^B$ in their DSGE-VAR framework with a conjugate prior. I am interested in the case of $T_0^V \neq T_0^B$ and when the prior is not conjugate. Thus, I analyze the case of increasing the degrees of freedom only of the Wishart prior, abstracting from unknown model dynamics (i.e., $\beta = 0$) so that T_0^B becomes irrelevant.

[45] As suggested by Rossi et al. (2005, ch. 2.12), I use an equivalent representation of the posterior for the numerical implementation: I use that $\mathcal{N}(\mathbf{0}, (\Sigma \otimes (X'X)^{-1})) \overset{D}{=} (\text{chol}(\Sigma) \otimes \text{chol}((X'X)^{-1}))\mathcal{N}(\mathbf{0}, I)$ and draw using the identity $(\text{chol}(\Sigma) \otimes \text{chol}((X'X)^{-1}))\,\text{vec}(Z) = vec(\text{chol}((X'X)^{-1}) Z \,\text{chol}(\Sigma)')$, with $\text{vec}(Z) \sim \mathcal{N}(\mathbf{0}, I)$.

The marginal likelihood of an *iid* sample of length T with $y_t \in \mathbb{R}^m$ is given by:

$$\begin{aligned}
p(y|T_0^V) &\equiv \int_0^\infty \left(\prod_{s=1}^T f(y_s|V)\right) \pi(V|T_0^V) dV^{-1} \\
&= \int_0^\infty (2\pi)^{-mT/2} |V|^{-T/2} e^{-\frac{T}{2}\operatorname{tr}(V^{-1}\hat{V}) - \frac{T_0^V}{2}\operatorname{tr}(V^{-1}V_0)} 2^{-mT_0^V/2} |V_0 T_0^V|^{T_0^V/2} \Gamma_m(T_0^V/2)^{-1} |V|^{-(T_0-m-1)/2} dV^{-1} \\
&= \pi^{-mT/2} \frac{|V_0 T_0^V|^{T_0^V/2} \Gamma_m((T+T_0^V)/2)}{|V_0 T_0^V + T\hat{V}|^{(T+T_0^V)/2} \Gamma_m(T_0^V/2)} \times \\
&\quad \int_0^\infty e^{-\frac{1}{2}\operatorname{tr}\left(V^{-1}\left(T\hat{V}+T_0^V V_0\right)\right)} 2^{-m(T+T_0^V)/2} |V_0 T_0^V + T\hat{V}|^{(T+T_0^V)/2} \Gamma_m((T+T_0^V)/2)^{-1} |V|^{-(T+T_0-2)/2} dV^{-1} \\
&= \pi^{-mT/2} \frac{|V_0 T_0^V|^{T_0^V/2} \Gamma_m((T+T_0^V)/2)}{|V_0 T_0^V + T\hat{V}|^{(T+T_0^V)/2} \Gamma_m(T_0^V/2)} \\
&= \pi^{-mT/2} |\hat{V}|^{-T/2} \frac{|\hat{V}^{-1} V_0|^{T_0^V/2}}{|\hat{V}^{-1} V_0 + I_m \frac{T}{T_0^V}|^{(T+T_0^V)/2}} \frac{(T_0^V)^{-mT/2} \Gamma_m((T+T_0^V)/2)}{\Gamma_m(T_0^V/2)} \qquad \text{(A.34)}
\end{aligned}$$

defining $\hat{V} \equiv \frac{1}{T}\sum_{s=1}^T y_s y'_s$ and using $\Gamma_m(T/2) = \pi^{m(m-1)/2} \prod_{j=1}^m \Gamma(\frac{1}{2}(T+1-j))$. It is straightforward to show via the first-order condition that the (log) data density is maximized by a DSGE model prior centered at $V_0 = \hat{V}$: the data rewards model fit.

To gain intuition, consider the scalar case $m = 1$. Abstracting from terms constant in T_0^V, the density can then be simplified to:

$$\ln p(y|T_0^V) = \kappa(V,T) - \frac{T}{2}\ln(T_0^V) + \frac{T_0^V}{2}\ln(\frac{V_0}{\hat{V}}) - \frac{T+T_0^V}{2}\ln(T_0^V \frac{V_0}{\hat{V}}) + \ln \frac{\Gamma\left(\frac{T+T_0^V}{2}\right)}{\ln\Gamma\left(\frac{T_0^V}{2}\right)}$$

The slope of the log data density in T_0^V is given by:

$$\frac{d\ln p(y|T_0^V)}{dT_0^V} = \frac{1}{2}\ln\left(\frac{T_0^V \frac{V_0}{\hat{V}}}{T_0^V \frac{V_0}{\hat{V}} + T}\right) + \frac{1}{2}\left(1 - \frac{V_0}{\hat{V}}\right)\frac{T}{T_0^V \frac{V_0}{\hat{V}} + T} + \frac{1}{2}\psi\left(\frac{T+T_0^V}{2}\right) - \frac{1}{2}\psi\left(\frac{T_0^V}{2}\right), \qquad \text{(A.35)}$$

where ψ is the digamma function, the derivative of the log gamma function. Part (a) of the following Lemma establishes that for $\frac{V_0}{\hat{V}}$ in an open neighborhood around unity, the slope of the log data density is strictly positive (at T that are multiples of 2). Hence, when the DSGE model V_0 fits the data well, an infinite prior weight on the DSGE model maximizes the fit. Parts (b)and (c) establish the counterpart that for a sufficiently bad fit so that $\frac{V_0}{\hat{V}}$ is far enough from unity, the slope of the log data density is negative in T_0^V. Thus, the optimal prior weight diverges.

Lemma 5. *Let $T = 2n, n \in \mathbb{N}_+$ and $T_0^V > 0$.*

(a) For $\frac{V_0}{V}$ in an open neighborhood around unity, $\frac{d}{dT_0^V}\ln p(y|T_0^V) > 0$.

(b) There exists a number $\underline{v} \in (0,1)$ such that for $\frac{V_0}{V} < \underline{v}$ $\frac{d}{dT_0^V}\ln p(y|T_0^V) < 0$.

(c) For $T > 2$, there exists a number $\bar{v} > 1$ such that for $\frac{V_0}{V} > \bar{v}$ $\frac{d}{dT_0^V}\ln p(y|T_0^V) < 0$.

Proof. Consider the three cases in the lemma separately:

(a) Let $V_0 = \hat{V}$. Note that under the assumption on T, the recurrence relation of the digamma function implies that

$$\psi\left(\frac{T+T_0^V}{2}\right) = \psi\left(\frac{T_0^V}{2}\right) + \sum_{s=0}^{\frac{T}{2}-1} \frac{1}{\frac{T_0^V}{2}+s}.$$

The slope (A.35) can therefore be written as:

$$\frac{1}{2}\ln\left(\frac{T_0^V}{T_0^V+T}\right) + \frac{1}{2}\sum_{s=0}^{\frac{T}{2}-1} \frac{1}{\frac{T_0^V}{2}+s}.$$

Note that for $x > 0$:

$$\frac{1}{2}\ln\left(\frac{x}{2+x}\right) + \frac{1}{x} > 0.$$

This inequality follows from a basic logarithm inequality: $-\log\left(\frac{x}{2+x}\right) = \log\left(1+\frac{2}{x}\right) < \frac{2}{x}$. Thus, $\frac{1}{2}\ln\left(\frac{x}{2+x}\right) + \frac{1}{x} > 0$.

The result on $\frac{d}{dT_0^V}\ln p(y|T_0^V)$ follows by induction for $V_0 = \hat{V}$. Let $T = 2 \Leftrightarrow n = 1$. Then the above inequality for $x = T_0^V$ implies the condition for $n = 1 \Leftrightarrow T = 2$.

Now assume that the condition holds for arbitrary $n \in \mathbb{N}_+$. Notice that

$$\frac{d}{dT_0^V}\ln p(y|T_0^V)\Big|_{T=2(n+1)} - \frac{d}{dT_0^V}\ln p(y|T_0^V)\Big|_{T=2n} = \frac{1}{2}\ln\left(\frac{2n+T_0}{2(n+1)+T_0^V}\right) + \frac{1}{2}\frac{2}{2n+T_0^V},$$

which is larger than zero by the above inequality. In addition, by assumption, $\frac{d\ln p(y|T_0^V)}{dT_0^V}\Big|_{T=2n} > 0$. It follows that $\frac{d}{dT_0^V}\ln p(y|T_0^V)\Big|_{T=2(n+1)} > 0$.

Since the assumption is true for $n = 1$, the desired result for $\frac{d}{dT_0^V}\ln p(y|T_0^V)$ follows for $V_0 = \hat{V}$ and any $n \in \mathbb{N}_+$ by induction.

Last, because $p(y|T_0^V)$ and its derivatives are continuous in V_0, the inequality holds for V_0 sufficiently close to $\hat{V}$.

(b) Fix T, T_0^V. Note that $\lim_{V_0/\hat{V}\searrow 0} \frac{d}{dT_0^V}\ln p(y|T_0^V) = -\infty$. Since the limit is $-\infty$, there exists a number $\underline{v}$ such that for $\frac{V_0}{\hat{V}} < \underline{v}$ $\frac{d}{dT_0^V}\ln p(y|T_0^V) < 0$ holds. Since, by (a), the inequality is not satisfied at $V_0 = \hat{V}$, it follows that $\underline{v} < 1$.

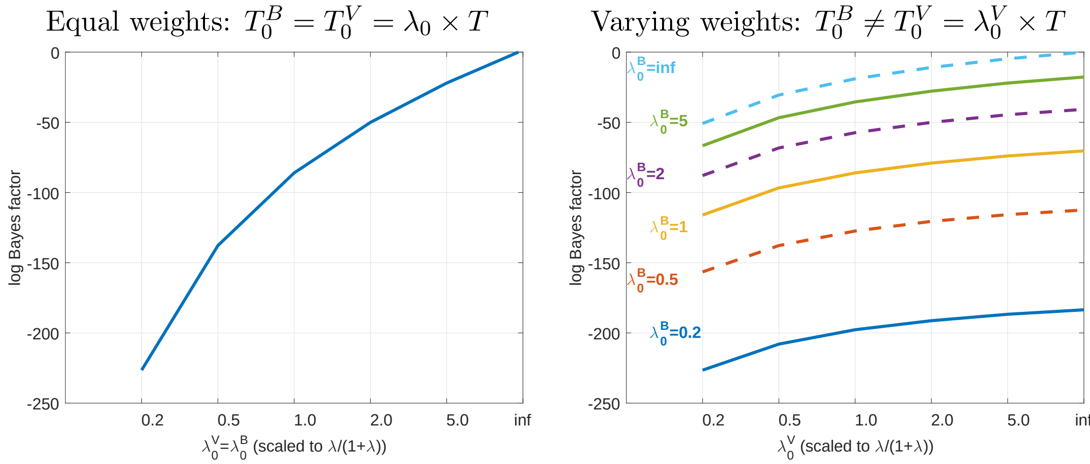
(c) Note that $\lim_{V_0/\hat{V}\to\infty} \frac{d}{dT_0^V}\ln p(y|T_0^V) = -\frac{T}{2T_0^V} + \frac{1}{2}\psi\left(\frac{T+T_0^V}{2}\right) - \frac{1}{2}\psi\left(\frac{T_0^V}{2}\right)$. Note also that $\psi\left(\frac{T+T_0^V}{2}\right) - \psi\left(\frac{T_0^V}{2}\right) \leq \frac{T}{2}\frac{2}{T_0}$ given the recurrence relation used in (a) and given that the sum in the recurrance relation has at most $\frac{T}{2}$ increments. These increments are smaller or equal to $\frac{2}{T_0}$. When $T > 2$, the equality is strict. Thus, $\lim_{V_0/\hat{V}\to\infty} \frac{d}{dT_0^V}\ln p(y|T_0^V) < 0$ for $T > 2$. By the definition of the limit, there exists some $\bar{v}$ such that the inequality holds for all $V_0 > \bar{v}\hat{V}$ and $T > 2$. By (a), $\bar{v} > 1$.

□

A.8.2 Numerical example

The logic behind the previous analytic results for the scalar case applies more widely: If the prior is sufficiently close to the data, increasing the prior precision increases the model fit. Here, I provide a numerical benchmark for the benchmark VAR specification.

Specifically, I abstract from uncertain DSGE (hyper-)parameters and fix the prior $\bar{B}_0^y, \bar{B}_0^z$ and $\bar{V}_0$ matrices so that the prior fit is perfect: I choose the prior to equal the posterior given the actual data. I then vary the prior precision T_0^B and T_0^V on a grid. As expected, the marginal likelihood is strictly increasing in both T_0^V and T_0^B and peaks at the limit point of $T_0^B = T_0^V \to \infty$.



In this numerical example, the prior is chosen to equal the posterior for the baseline narrative DSGE-VAR. Thus, the prior fits the data as well as possible. The figures show that increasing the weights $T_0{}^B$, $T_0{}^V$ strictly increases the model fit, which is measured via the marginal likelihood. To give some context, the number of prior observations is expressed as $T_0 = \lambda_0 \times T$ (i.e., relative to the empirical sample size). The marginal likelihood is strictly increasing in both the dimension of "dynamics" via the number of dummy observations on the coefficient matrix and the dimension of "identification" via the number of dummy observations on the covariance matrix.

Figure A.1: Narrative DSGE-VAR marginal likelihood with fixed hyperparameters when prior is set to equal the posterior

A.9 Comparing estimators for the marginal likelihood

Besides the nested Chib (1995) likelihood estimator within Geweke (1999), I also use the iterative bridge sampler (Meng and Wong, 1996) and importance sampling as described in Geweke (2005, ch. 8.2). The latter exploits that the likelihood itself in my application does not depend on the models that I am comparing. The Bayes factor can then be written as:

$$\begin{aligned} BF(\tau,\tau') &= \frac{\int_\Theta \int_{\mathbb{R}^{mp\times k+m_z+1}} \int_{V^{-1}\in PD(m+m_z)} f(y,z|B,V^{-1})\pi(B,V^{-1}|\theta;\tau)\pi(\theta)dBdV^{-1}d\theta}{\int_\Theta \int_{\mathbb{R}^{mp\times k+m_z+1}} \int_{V^{-1}\in PD(m+m_z)} f(y,z|B,V^{-1})\pi(B,V^{-1}|\theta;\tau')\pi(\theta)dBdV^{-1}d\theta} \\ &\equiv \int_\Theta \int_{\mathbb{R}^{mp\times k+m_z+1}} \int_{V^{-1}\in PD(m+m_z)} \frac{\pi(B,V^{-1}|\theta;\tau)}{\pi(B,V^{-1}|\theta;\tau')}\pi(\theta,B,V^{-1}|y,z;\tau')dBdV^{-1}d\theta, \end{aligned}$$

where

$$\pi(\theta,B,V^{-1}|y,z;\tau') \equiv \frac{f(y,z|B,V^{-1})\pi(B,V^{-1}|\theta;\tau')\pi(\theta)}{\int_\Theta \int_{\mathbb{R}^{mp\times k+m_z+1}} \int_{V^{-1}\in PD(m+m_z)} f(y,z|B,V^{-1})\pi(B,V^{-1}|\theta;\tau')\pi(\theta)dBdV^{-1}d\theta}$$

is the posterior over the parameters under τ'. I use the following simulation-consistent estimator for the Bayes factor:

$$\widehat{BF}(\tau, \tau') = N^{-1} \sum_{i=1}^{N} \frac{\pi(B_{(i)}, V_{(i)}^{-1} | \theta_{(i)}; \tau)}{\pi(B_{(i)}, V_{(i)}^{-1} | \theta_{(i)}; \tau')},$$

where $B_{(i)}, V_{(i)}^{-1}, \theta_{(i)}$ are draws from the posterior given τ'.

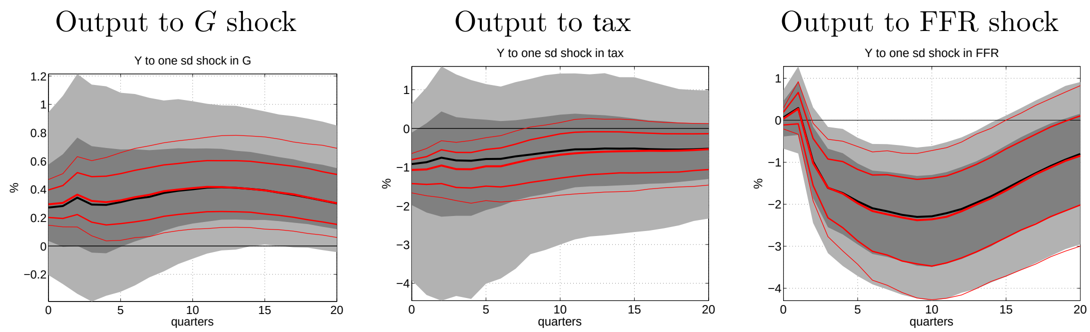
$T_0^B = T_0^V$	Chib-Geweke (Baseline)	Importance Sampling	Meng-Wong
0.2×T	-32.3	-63.0	-32.9
0.5×T	-12.9	-25.6	-15.8
1.0×T	-4.1	-7.0	-4.9
1.5×T	-3.9	-1.7	-1.2
2.0×T	0.0	-0.4	-0.0
2.5×T	-8.7	0.0	0.0
3.0×T	-11.9	-0.5	-0.6
4.0×T	-15.8	-2.5	-3.0

Shown are the (natural) log of the Bayes factors relative to the best model for different estimators as a function of the overall prior precision T_0. See the text for details about the different estimators.

Table A.1: Comparing marginal likelihood estimators in the baseline model

B Sampling properties of VAR estimator

The estimated credible sets are wide – wider than they would be if uncertainty about Γ were ignored. Rather than drawing Γ as part of V from the posterior, one can compute $\hat{\Gamma} = \frac{1}{T}(Y - X_y B_y)'(Z - X_z B_z)$ conditional on the estimated coefficients. Figure B.2 compares the posterior uncertainty over the output response to different shocks for both sampling schemes.



Comparing the estimated credible sets in gray to credible sets that abstract from sampling error in the estimated covariance produce much tighter estimates when instruments used to identify shocks have only few non-zero observations as is the case for spending and tax shocks. Note: Shown are the pointwise median and 68% and 90% posterior credible sets of the SUR-sampling scheme (black and gray) and when the point estimate $\hat{\Gamma} = \frac{1}{T}(Y - X_y B_y)'(Z - X_z B_z)$ is used for Γ (red). Results based on lower Cholesky factorization of $S_1 S_1'$.

Figure B.2: Effects of policy shocks on output: comparing credible sets 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%.

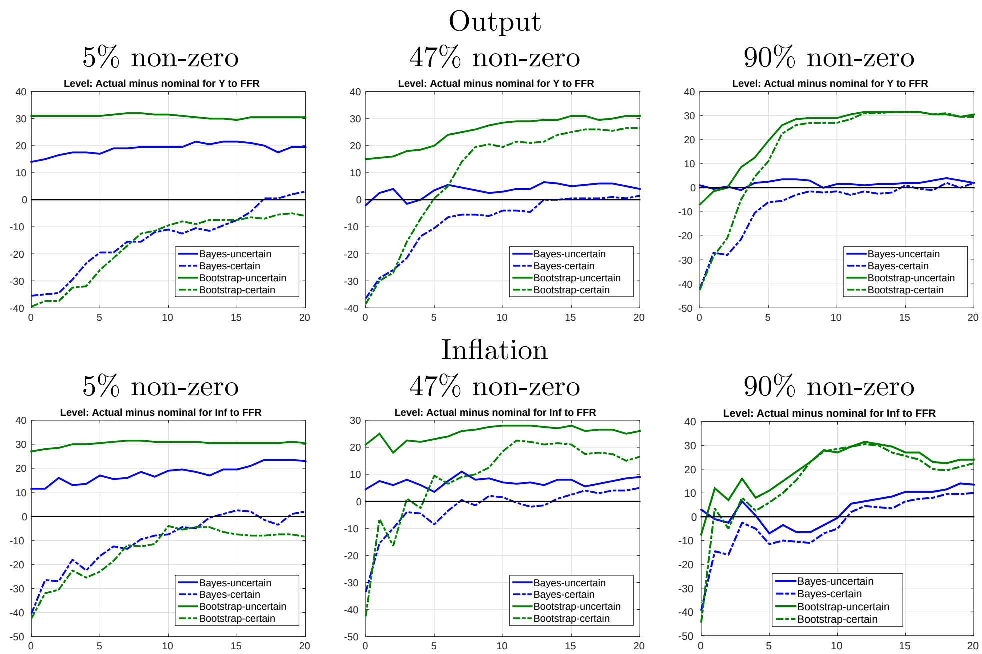
For each dataset m, I compute the pointwise posterior credible set using my Bayesian procedure, with and without conditioning on the observed covariance between instruments and VAR forecast errors. Similarly, I use the wild bootstrap proposed in Mertens and Ravn (2013) to conduct frequentist inference.[46] Each procedure yields an estimate of the true IRF $\{I_{h,m}\}_{h=0}^{H}$ for each point OLS estimate and a pointwise credible set for horizons up to H: $\{\hat{\mathcal{C}}(\alpha)_{h,m}^{j}\}_{h=0}^{H} = \{[\underline{c}(\alpha)_{h,m}^{j}, \overline{c}(\alpha)_{h,m}^{j}]\}_{h=0}^{H}$. The superscript j indexes the different methods. $1 - \alpha$ is the nominal size of the credible set: A fraction α of draws from the posterior or the bootstrapped distribution for $I_{h,m}$ should lie under $\underline{c}(\alpha)_{h,m}^{j}$ or above $\overline{c}(\alpha)_{h,m}^{j}$.

To assess the actual level of the credible sets of method j, I compute for each (h, m) whether the truth at horizon h, $I_{h,m}$ lies inside the pointwise credible set: $\mathbf{1}\{I_{h,m} \in \hat{\mathcal{C}}(\alpha)_{h,m}^{j}\}$. The actual coverage probability is then estimated as:

[46] Appendix B describes the algorithm.

$$\hat{\alpha}_h^j = \frac{1}{N_{MC}} \sum_{m=1}^{N_{MC}} \mathbf{1}\{I_{h,m} \in \hat{\mathcal{C}}(\alpha)_{h,m}^j\}$$

Figure B.3 plots the deviation of the actual coverage probability for model j at horizon h from the nominal coverage probability: $\hat{\alpha}_h^j - \alpha$ 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 credible sets substantially at short horizons ("Bayes – certain", BaC and "Bootstrap – certain," BoC), while the Bayesian procedure that 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.[47] However, when only one in 20 observations for the instrument is non-zero, the actual level exceeds the nominal level by around 10 percentage points.



With enough observations on narrative instruments, the proposed estimator (shown in solid blue) has good classical coverage probabilities. With few observations, the estimator seems to be too conservative, whereas estimators neglecting uncertainty in the covariance matrix provide misleading tight confidence bands.

Figure B.3: Monte Carlo analysis of actual minus nominal coverage probability $\hat{\alpha}_h^j - \alpha$

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 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.

[47] The actual level depends slightly on the variable under consideration.

Frequentist inference

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

1. For $t = 1, \ldots, T$, draw $\{e_1^b, \ldots, e_T^b\}$, where $e_t^b \sim iid$ with $\Pr\{e_t^b = 1\} = \Pr\{e_t^b = -1\} = 0.5$.
2. Construct the artificial data for Y_t^b. In a VAR of lag length p, build $Y_t^b, t > p$ as:
   - For $t = 1, \ldots, p$ set $Y_t^b = Y_t$.
   - For $t = p+1, \ldots, T$ construct recursively $Y_t^b = \sum_{j=1}^{p} \hat{B}_j Y_{t-j}^b + e_t^b \hat{u}_t$.
3. Construct the artificial data for the narrative instrument:
   - For $t = 1, \ldots, T$ construct recursively $z_t^b = e_t^b z_t^b$.

C Data and additional results

C.1 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:

$$y_t = \frac{(\text{Nominal GDP: NIPA Table 1.1.5Q, Line 1})_t}{(\text{Population above 16: FRED CNP16OV})_t \times (\text{GDP deflator: NIPA Table 1.1.9Q, Line 1})_t}$$

$$c_t = \frac{(\text{Nominal PCE on nondurables and services: NIPA Table 1.1.5Q, Lines 5+6})_t}{(\text{Population above 16: FRED CNP16OV})_t \times (\text{GDP deflator: NIPA Table 1.1.9Q, Line 1})_t}$$

$$i_t = \frac{(\text{Durables PCE and fixed investment: NIPA Table 1.1.5Q, Lines 4 + 8})_t}{(\text{Population above 16: FRED CNP16OV})_t \times (\text{GDP deflator: NIPA Table 1.1.9Q, Line 1})_t}$$

$$\pi_t = \Delta \ln(\text{GDP deflator: NIPA Table 1.1.9Q, Line 1})_t$$

$$r_t = \begin{cases} \frac{1}{4}(\text{Effective Federal Funds Rate: FRED FEDFUNDS})_t & t \geq (1954\text{:Q3} \\ \frac{1}{4}(\text{3-Month Treasury Bill: FRED TB3MS})_t & (\text{else}. \end{cases}$$

$$n_t = \frac{(\text{Nonfarm business hours worked: BLS PRS85006033})_t}{(\text{Population above 16: FRED CNP16OV})_t}$$

$$w_t = \frac{(\text{Nonfarm business hourly compensation: BLS PRS85006103})_t}{(\text{GDP deflator: NIPA Table 1.1.9Q, Line 1})_t}$$

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 are computed following Leeper et al. (2010), except for adding state and local govern-ments (superscript "s&l") to the federal government account (superscript "f"), similar to Fernandez-Villaverde et al. (2015). Since in the real world

$$\tau_t^c = \frac{(\text{Production \& imports taxes: Table 3.2, Line 4})_t{}^f + (\text{Sales taxes})_t{}^{s\&l}}{((\text{Durables PCE})_t + c_t) \times (\text{GDP deflator})_t - (\text{production \& imports taxes})_t^f - (\text{Sales taxes})_t^{s\&l}}$$

$$\tau_t^p = \frac{(\text{Personal current taxes})_t}{\frac{1}{2}(\text{Proprietors' income})_t + (\text{wage income})_t + (\text{wage supplements})_t + (\text{capital income})_t}$$

$$\tau_t^n = \frac{\tau_t^p(\frac{1}{2}(\text{Proprietors' income})_t + (\text{wage income})_t + (\text{wage supplements})_t) + (\text{wage taxes})_t^f}{(\text{wage income})_t + (\text{wage supplements})_t + (\text{wage taxes})_t^f + \frac{1}{2}(\text{Proprietors' income})_t}$$

$$\tau_t^k = \frac{\tau_t^p(\text{capital income})_t + (\text{corporate taxes})_t^f + (\text{corporate taxes})_t^{s\&l}}{(\text{Capital income})_t + (\text{Property taxes})_t^{s\&l}}$$

where the following NIPA sources were used:

- (Federal) production and 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.

$$(\text{rev})_t^c = \frac{\tau_t^c \times (c_t - (\text{Taxes on production and imports})_t^f - (\text{Sales taxes})_t^{s\&l}}{((\text{Population above 16})_t \times ((\text{GDP deflator})_t}$$

$$(\text{rev})_t^n = \tau_t^n \times ((\text{wage income})_t + (\text{wage supplements})_t + (\text{wage taxes})_t^f + \frac{1}{2}(\text{Proprietors' income})_t)$$

$$(\text{rev})_t^k = \tau_t^k \times ((\text{Capital income})_t + (\text{Property taxes})_t^{s\&l})$$

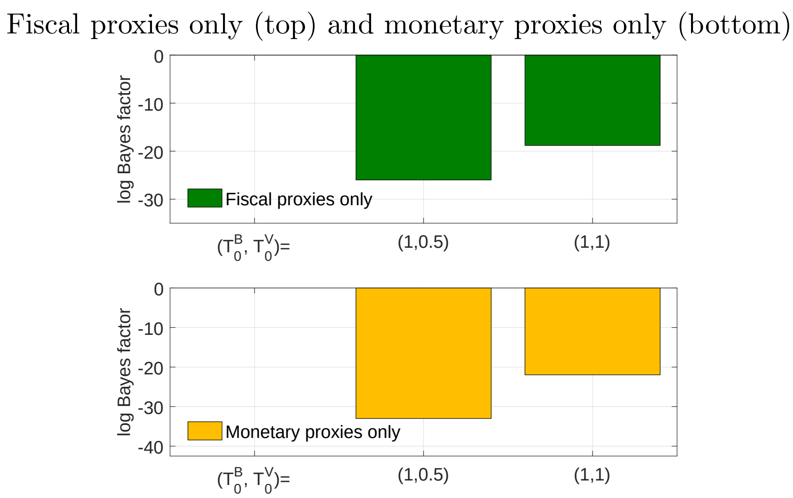
I construct government debt as the cumulative net borrowing of the consolidated NIPA government sector and adjust the level of debt to match the value of consolidated government FoF debt at par value in 1950:Q1. A minor complication arises as federal net purchases of nonproduced assets (NIPA

Table 3.2Q, Line 43) is missing prior to 1959:Q3. 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 lead 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 were revised up. The revision covers all quarters since 2004 but not before, implying a jump in the debt time series.[48] 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 are combined with data from the web appendices of Romer and Romer (2004), Ramey (2011), and Mertens and Ravn (2013) on narrative shock measures.

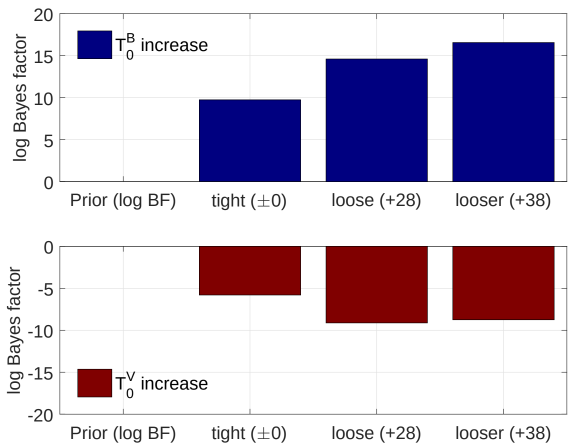
C.2 Additional results



The two panels show Bayes factors obtained by increasing the weight on the DSGE model covariance structure T_0^V, holding the weight on dynamics, T_0^B, fixed. Here, I use only a subset of the instruments in the estimation: Only fiscal proxies in the top panel and only monetary shock proxies in the bottom. Independent of the prior the main result carries through: The model fit is decreasing in the weight on the covariance structure.

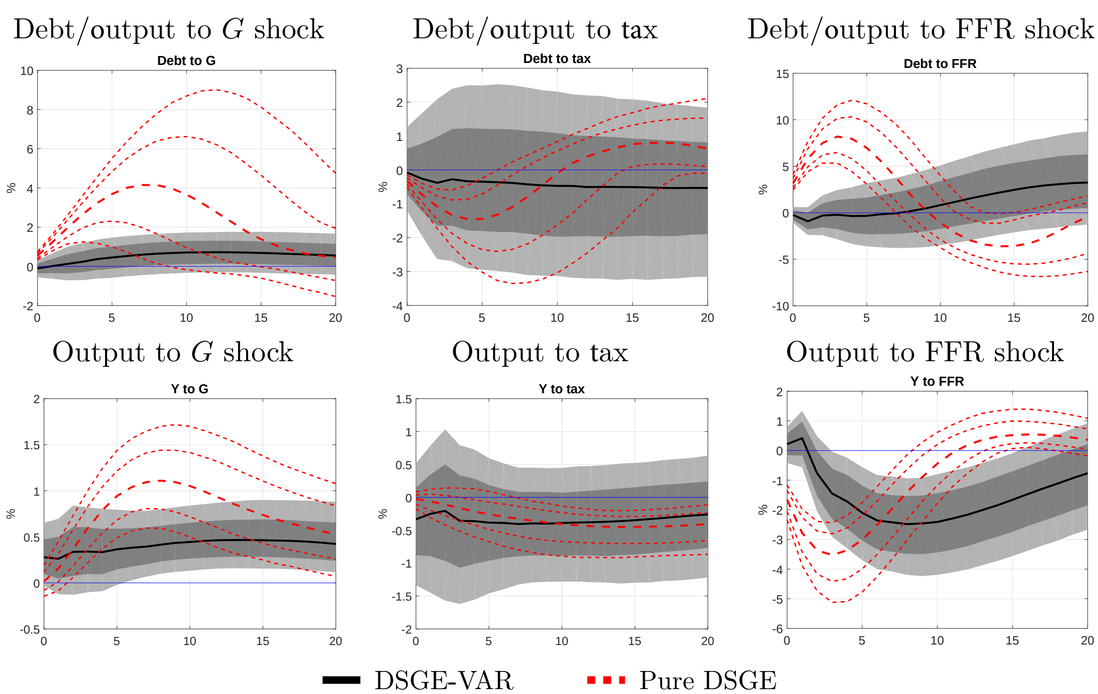
Figure C.4: Effect on Bayes factors of increasing T_0^V by $0.5T$ – separately with fiscal and monetary proxies

[48] 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).



The two panels show Bayes factors obtained by increasing the weight on model dynamics T_0^B (top panel) and the model covariance structure ${T_0}^V$ separately, starting from ${T_0}^B = {T_0}^V = 2 \times T$ for three different prior precision parameters. The "tight" parametrization is the benchmark and uses a prior standard deviation for the relative standard deviations of 0.1 and a prior standard deviation of 0.25 for the loadings. The "loose" prior uses a common standard deviation of 0.25. Last, the "looser" prior uses a common standard deviation of 0.5. As indicated in parentheses, the looser prior has the best overall model fit (highest Bayes factor) relative to the tight prior that is the benchmark. However, independent of the prior, the main result carries through: The model fit is increasing in the weight on model dynamics and decreasing in the weight on the covariance structure.

Figure C.5: Effect on Bayes factors of increasing ${T_0}^B$ (top) or ${T_0}^V$ (bottom) by $0.5T$, \ one at a time for different priors



Note: Shown are the pointwise median and 68% and 90% posterior credible sets. Results based on lower Cholesky factorization of $S_1 S_1'$.

Figure C.6: Response of the debt-to-output ratio to the identified policy shocks

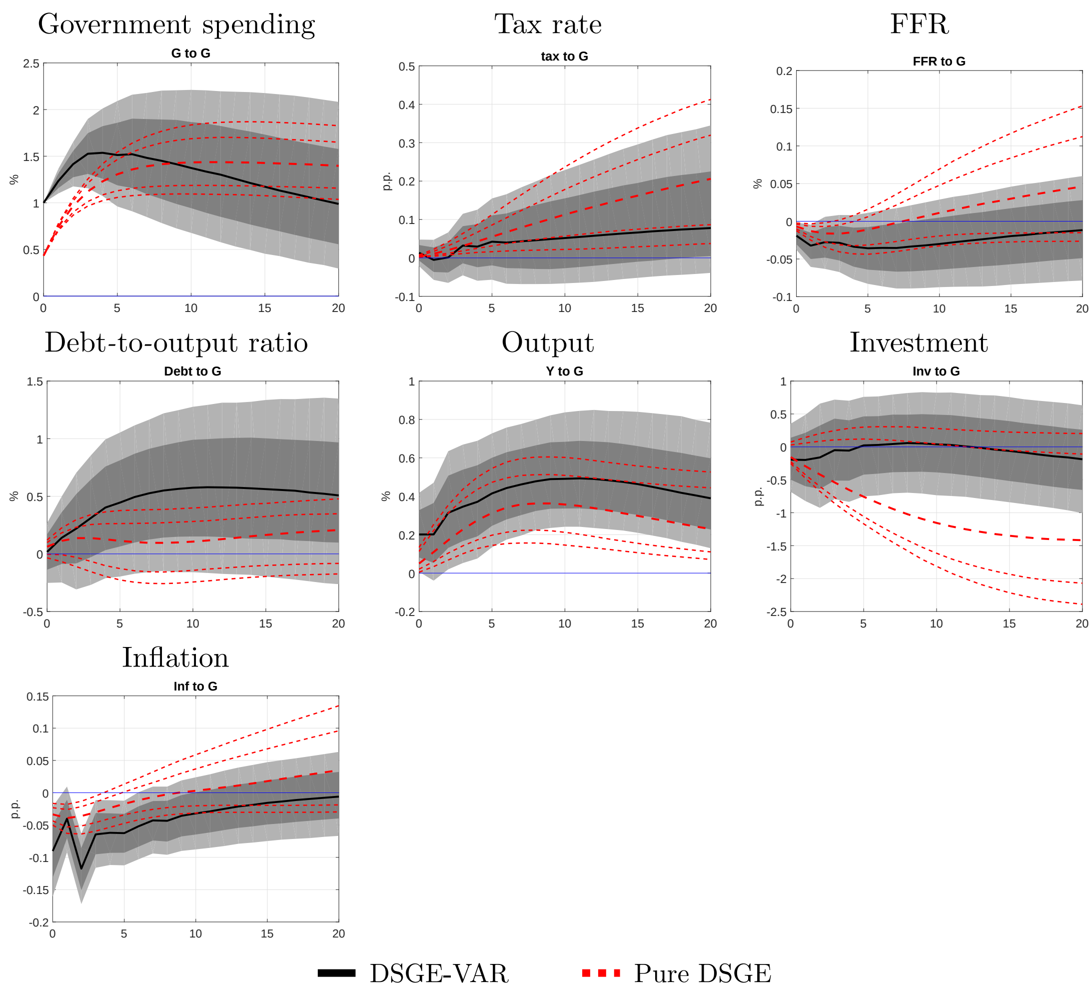


Figure C.7: Full set of responses to spending policy shock with best-fitting model $T_0 = 2 \times T$

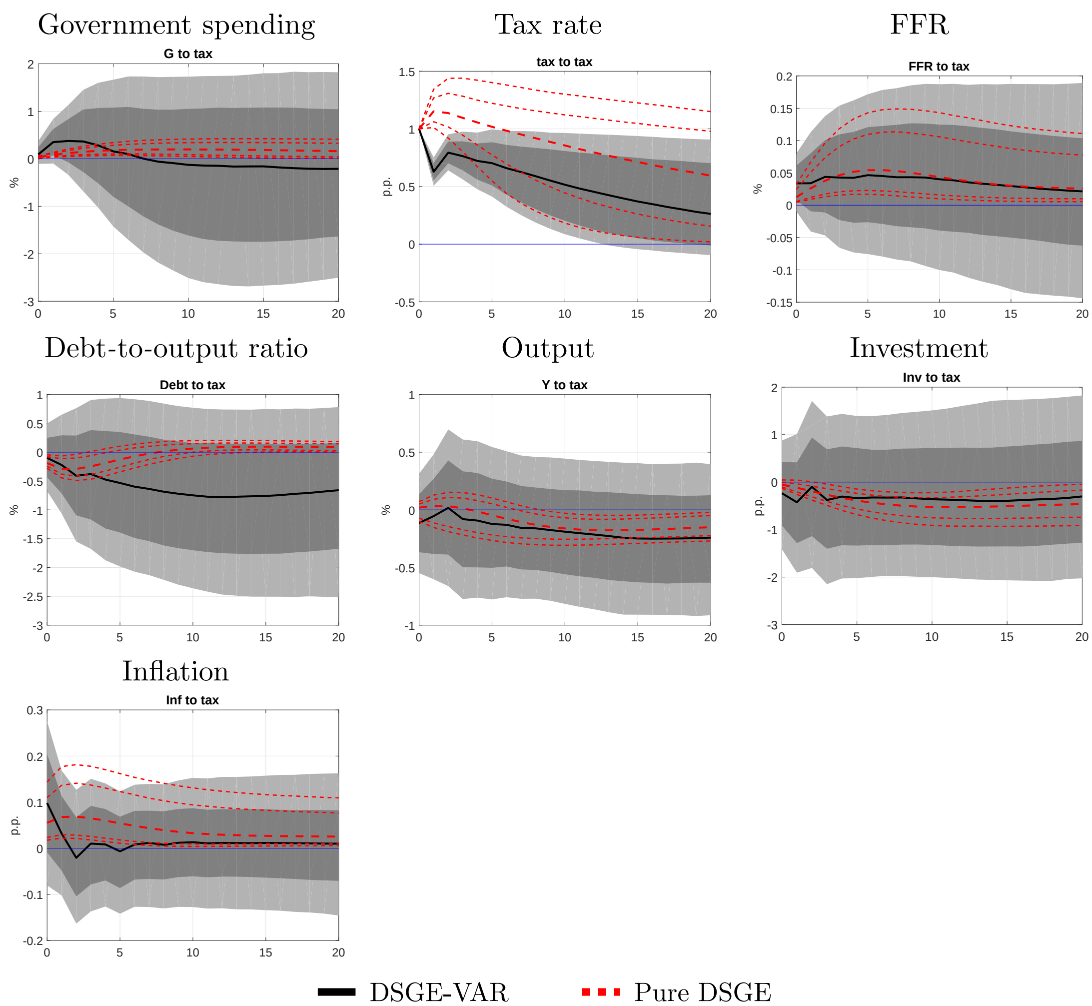


Figure C.8: Full set of responses to a tax rate shock with weak prior $T_0 = 2 \times T$

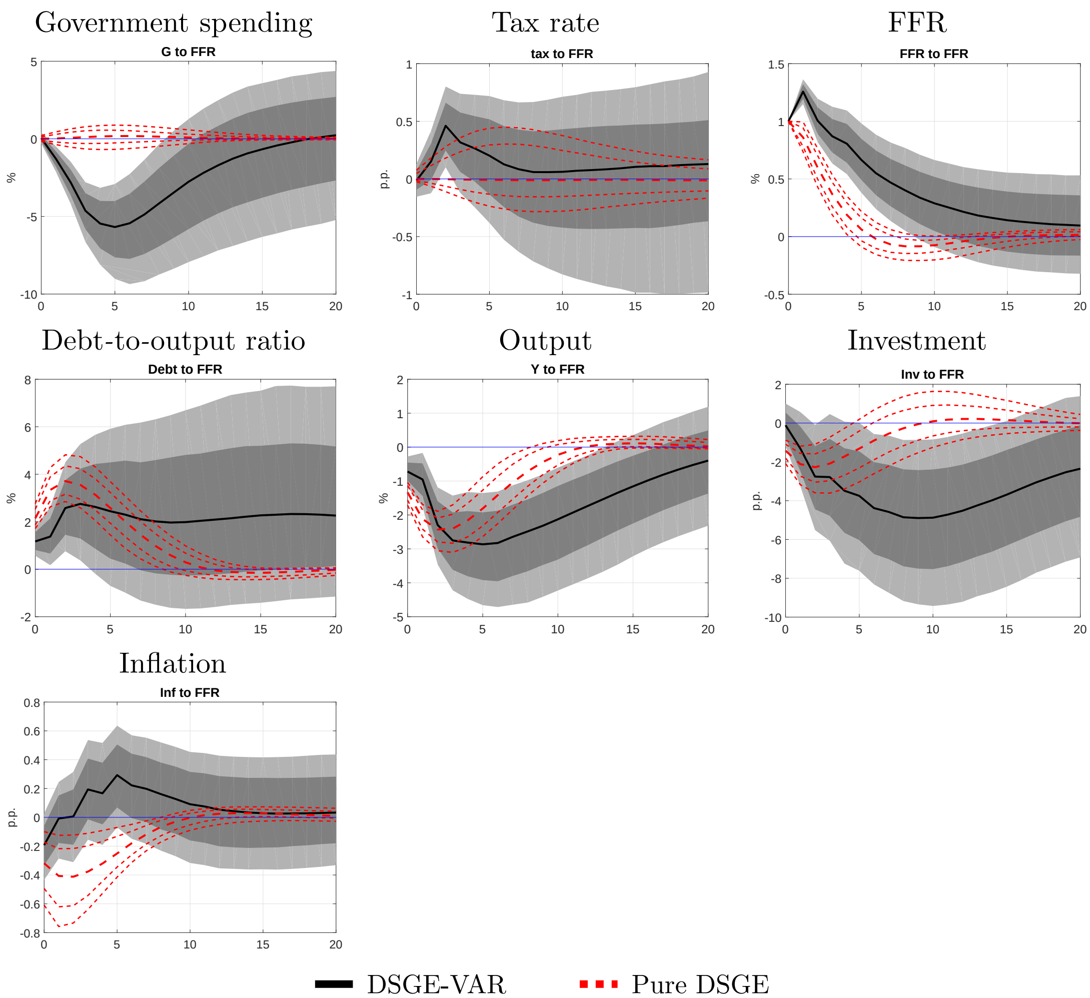
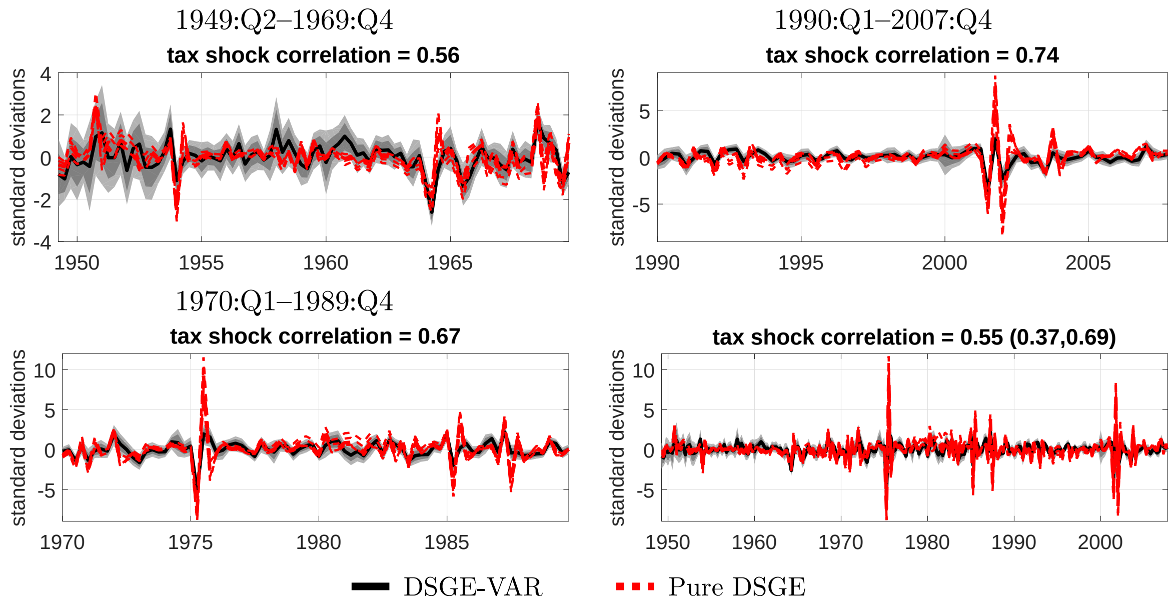
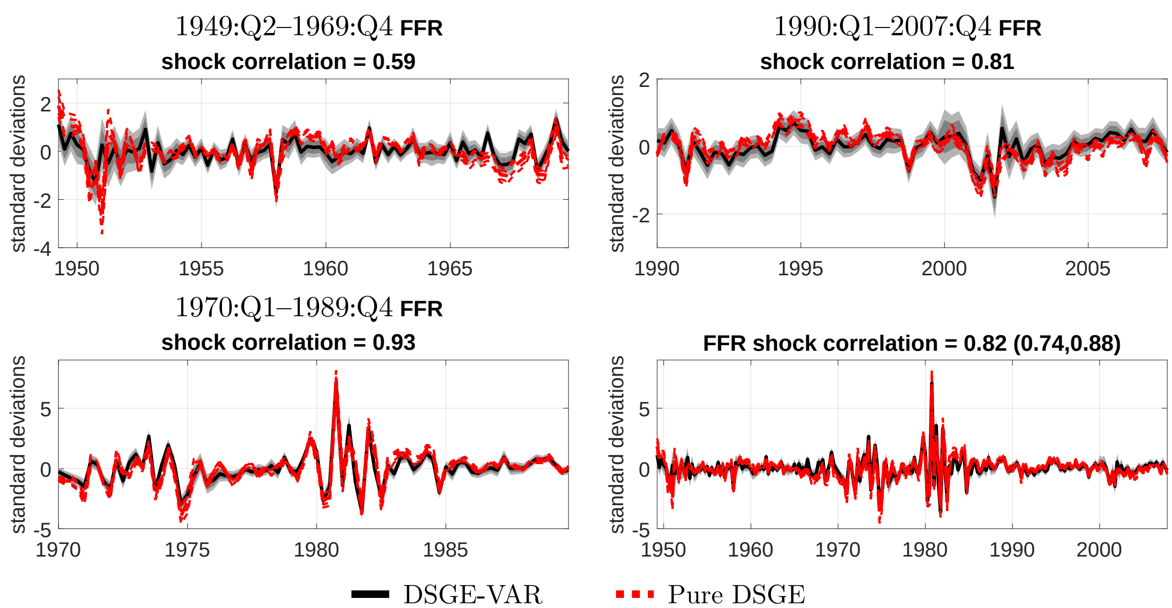


Figure C.9: Full set of responses to a FFR shock with best-fitting model $T_0 = 2 \times T$



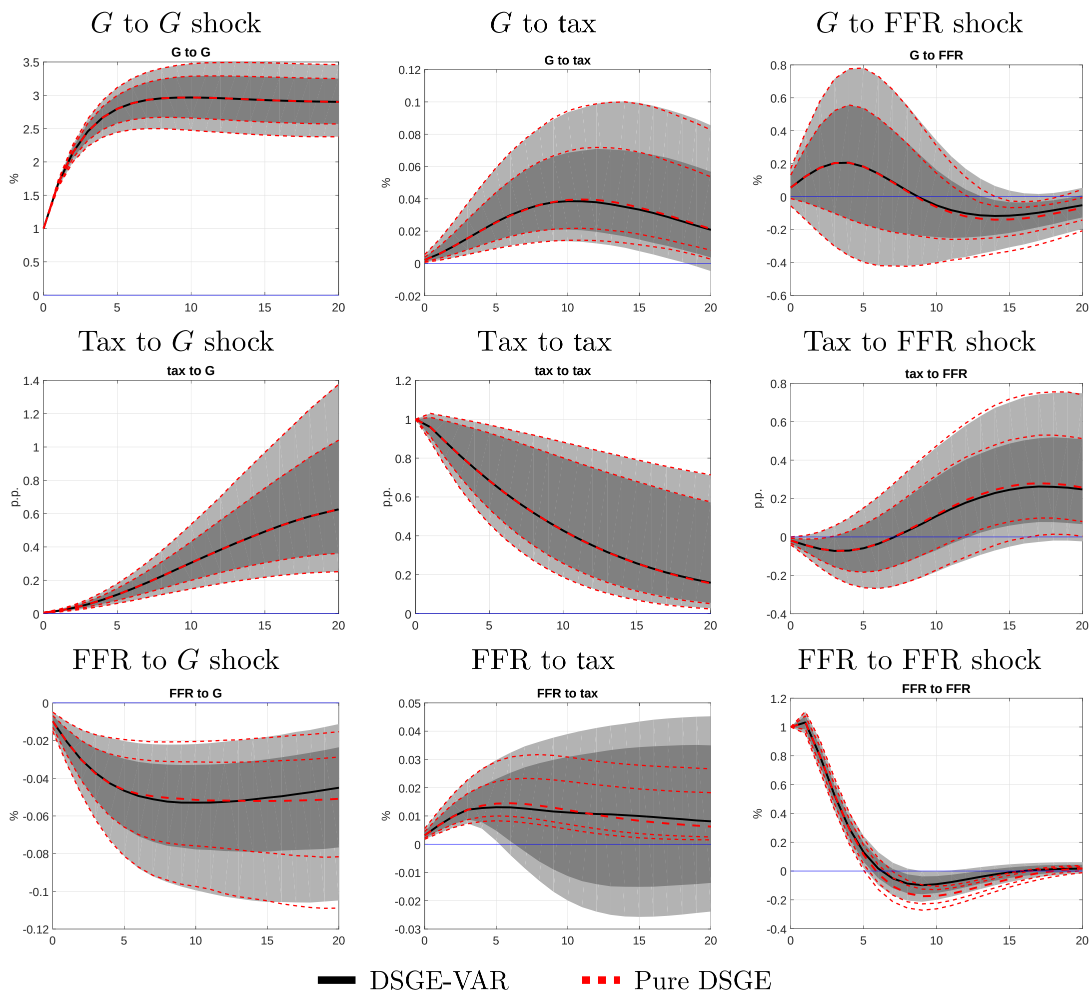
Estimated historical shocks. Note: Shown are the pointwise median and 68% and 90% posterior credible sets of shocks along with the median, 5th, and 95th percentiles of the shock correlations for the full sample. Results based on lower Cholesky factorization of $S_1 S_1'$.

Figure C.10: Historical tax shocks



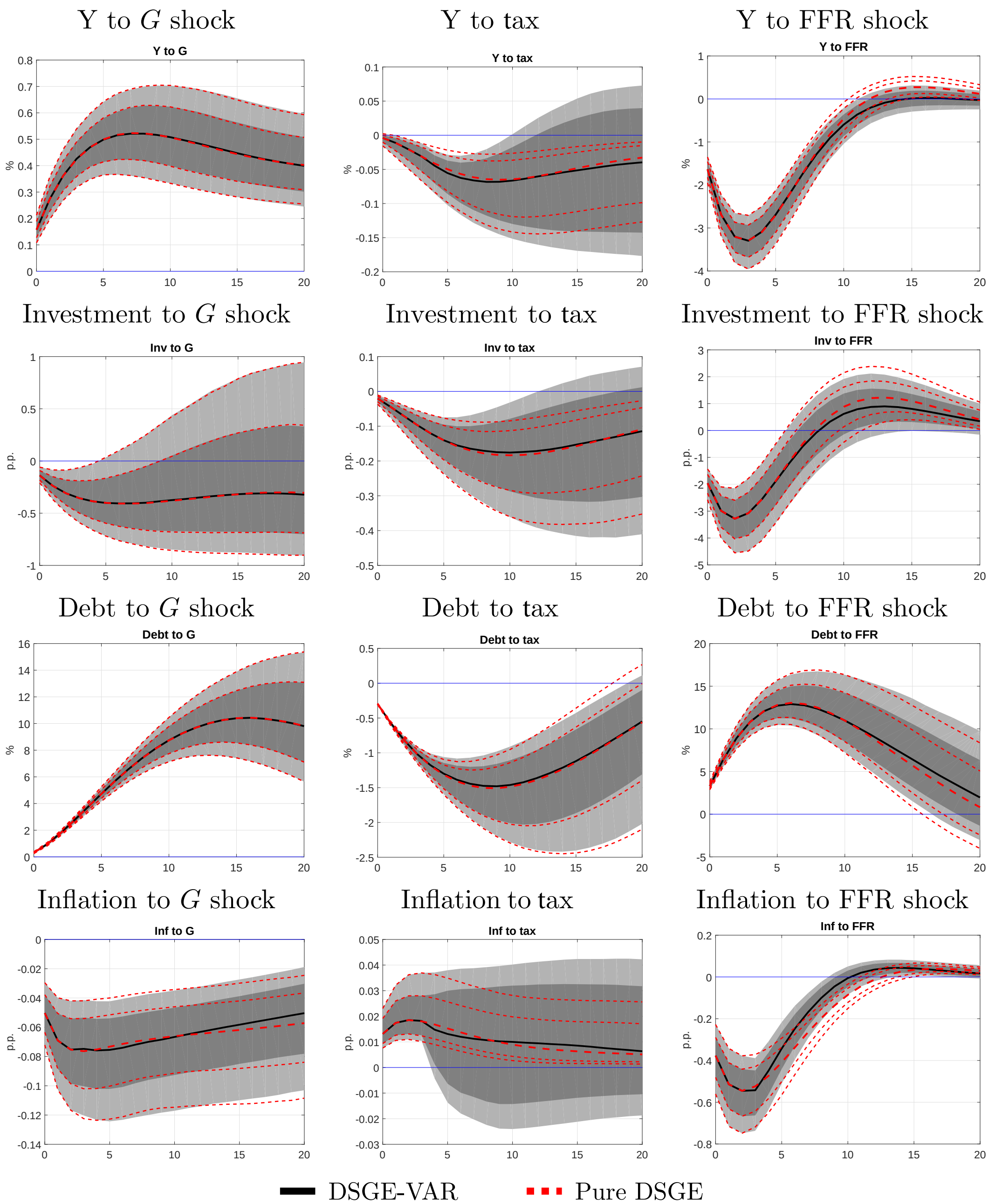
Estimated historical shocks. Note: Shown are the pointwise median and 68% and 90% posterior credible sets of shocks along with the median, 5th, and 95th percentiles of the shock correlations for the full sample. Results based on lower Cholesky factorization of $S_1 S_1'$.

Figure C.11: Historical FFR shocks



The IRFs implied by the posterior over θ largely coincide independent of whether I use the VAR(4) approximation or the state-space representation of the DSGE model. Note: Shown are the pointwise median and 68% and 90% posterior credible sets. Results based on lower Cholesky factorization of $S_1 S_1'$.

Figure C.12: Responses of policy instruments: DSGE model vs DSGE-VAR with dogmatic prior ($T_0^V = T_0^B \nearrow \infty$)



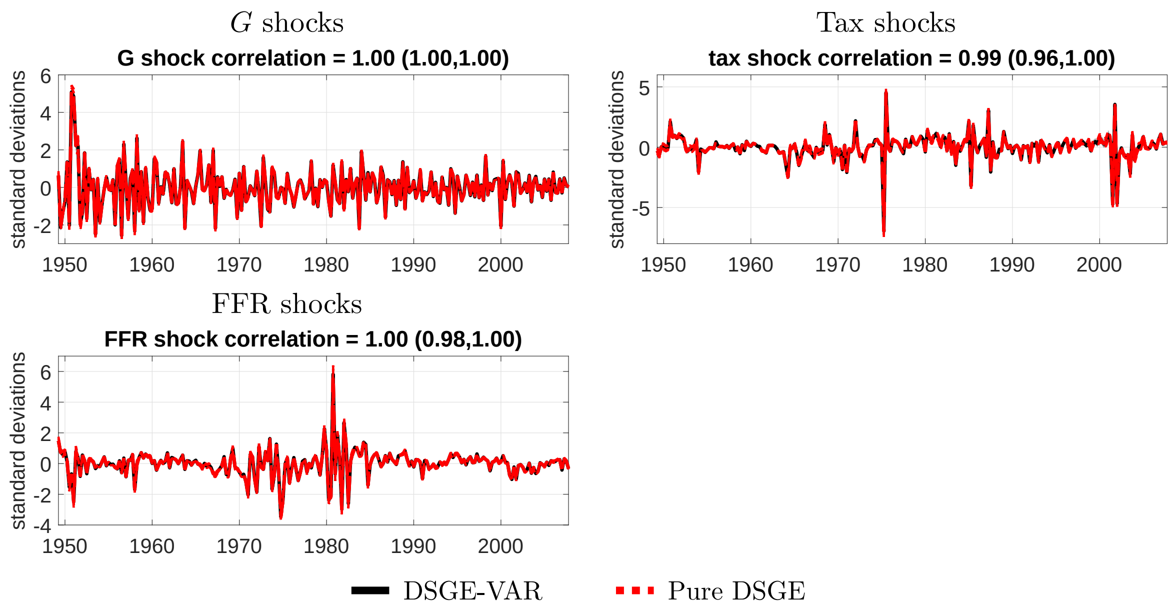
The IRFs implied by the posterior over θ largely coincide independent of whether I use the VAR(4) approximation or the state-space representation of the DSGE model. Note: Shown are the pointwise median and 68% and 90% posterior credible sets. Results based on lower Cholesky factorization of $S_1 S_1'$.

Figure C.13: Responses of output, investment, and inflation: DSGE model vs DSGE-VAR with dogmatic prior ($T_0^V = T_0^B \nearrow \infty$)

Parameter	Prior mean	(SD)	$T_0 = \frac{1}{5} \times T$ Mean	(SD)	$T_0 = 2 \times T$ Mean	(SD)	$T_0 = \infty$ Mean	(SD)
st.dev.-TFP	1.000	(2.000)	1.102	(0.214)	1.323	(0.242)	1.749	(0.184)
AR(1)-TFP	0.500	(0.200)	0.308	(0.150)	0.201	(0.112)	0.427	(0.087)
st.dev.-Transf.	1.000	(2.000)	0.715	(0.114)	1.743	(0.193)	1.773	(0.089)
st.dev.-G	1.000	(2.000)	0.783	(0.077)	1.171	(0.064)	1.420	(0.059)
AR(1)-G	0.500	(0.200)	0.895	(0.070)	0.991	(0.005)	0.998	(0.001)
st.dev.-Tax	0.500	(0.100)	0.301	(0.023)	0.304	(0.015)	0.332	(0.013)
AR(1)-Tax	0.500	(0.200)	0.560	(0.187)	0.225	(0.125)	0.034	(0.020)
st.dev.-qs	1.000	(2.000)	0.536	(0.117)	0.815	(0.161)	0.910	(0.078)
AR(1)-qs	0.500	(0.200)	0.794	(0.136)	0.444	(0.163)	0.457	(0.061)
st.dev.-FFR	1.000	(2.000)	0.222	(0.020)	0.237	(0.013)	0.258	(0.014)
AR(1)-FFR	0.500	(0.200)	0.301	(0.129)	0.159	(0.073)	0.320	(0.057)
st.dev.-Infl.	1.000	(2.000)	0.165	(0.033)	0.175	(0.040)	0.249	(0.018)
AR(1)-Infl.	0.500	(0.200)	0.323	(0.174)	0.345	(0.244)	0.121	(0.078)
Adj. cost	4.000	(1.500)	4.928	(1.373)	6.463	(1.147)	6.895	(0.931)
Util. cost	0.500	(0.150)	0.423	(0.119)	0.467	(0.159)	0.616	(0.066)
Fixed cost	1.250	(0.125)	1.376	(0.105)	1.355	(0.107)	1.302	(0.099)
Habit	0.700	(0.100)	0.752	(0.056)	0.717	(0.057)	0.715	(0.035)
Labor supply ela.	2.000	(0.750)	2.034	(0.662)	3.074	(0.890)	7.503	(1.310)
Calvo prices	0.500	(0.100)	0.689	(0.061)	0.735	(0.068)	0.806	(0.030)
Index. prices	0.500	(0.150)	0.419	(0.160)	0.290	(0.110)	0.393	(0.125)
Calvo wages	0.500	(0.100)	0.687	(0.089)	0.669	(0.078)	0.797	(0.036)
Index. wages	0.500	(0.150)	0.493	(0.150)	0.527	(0.142)	0.616	(0.127)
Taylor-Infl.	1.500	(0.250)	1.501	(0.189)	1.362	(0.300)	1.523	(0.142)
Taylor-GDP	0.125	(0.050)	0.155	(0.039)	0.112	(0.068)	0.096	(0.026)
smoothing-FFR	0.750	(0.100)	0.851	(0.031)	0.828	(0.027)	0.839	(0.021)
G-to-GDP	0.500	(0.500)	0.452	(0.459)	0.622	(0.466)	0.144	(0.141)
G-to-Debt	1.000	(0.500)	0.716	(0.402)	0.410	(0.189)	0.010	(0.006)
smoothing-G	0.500	(0.200)	0.956	(0.044)	0.797	(0.052)	0.695	(0.035)
Tax-to-GDP	1.000	(0.500)	0.601	(0.491)	0.581	(0.622)	0.467	(0.612)
Tax-to-Debt	1.000	(0.500)	0.291	(0.193)	0.327	(0.208)	0.068	(0.100)
smoothing-Tax	0.500	(0.200)	0.952	(0.062)	0.944	(0.071)	0.930	(0.045)
Transf.-to-GDP	0.000	(0.500)	-0.022	(0.354)	0.249	(0.274)	0.888	(0.390)
Transf.-to-Debt	1.000	(0.500)	0.979	(0.501)	1.276	(0.287)	0.243	(0.132)
smoothing-Transf.	0.500	(0.200)	0.910	(0.100)	0.251	(0.113)	0.938	(0.027)
Rel. st. dev. IV-G	1.000	(0.100)	1.034	(0.090)	0.846	(0.053)	0.746	(0.041)
Rel. st. dev. IV-Tax	1.000	(0.100)	0.858	(0.066)	0.801	(0.046)	0.751	(0.039)
Rel. st. dev. IV-FFR	1.000	(0.100)	1.196	(0.123)	2.077	(0.137)	2.169	(0.140)
Loading IV-G	1.000	(0.500)	0.550	(0.118)	0.315	(0.044)	0.238	(0.030)
Loading IV-Tax	1.000	(0.500)	0.440	(0.084)	0.288	(0.042)	0.232	(0.030)
Loading IV-FFR	1.000	(0.500)	2.705	(0.386)	1.941	(0.251)	1.626	(0.195)

Note: Shown are the prior and posterior mean and standard deviation for the DSGE model parameters. Parameters that determine the shock processes are shown in the top, policy rule estimates in the middle, and the observation equations for instruments in the bottom. Overall, the posteriors differ significantly from the priors, consistent with identification of the parameters.

Table C.2: DSGE-VAR model parameter estimates with varying priors



The historical shocks implied by the posterior over θ largely coincide independent of whether I use the VAR(4) approximation or the state-space representation of the DSGE model. Note: Shown are the pointwise median and 68% and 90% posterior credible sets of shocks along with the median, 5th, and 95th percentiles of the shock correlations. Results based on lower Cholesky factorization of $S_1 S_1'$.

Figure C.14: Historical policy shocks: DSGE model vs DSGE-VAR with dogmatic prior ($T_0^V = T_0^B \nearrow \infty$)

C.2.1 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?

Here, I report results from testing this assumption parametrically: I compare the model fit under different assumptions about the narrative instruments by relaxing the assumption that each instrument loads on only one structural shock. To do so, I re-estimate the pure DSGE model, including one narrative shock at a time with different restrictions in the observation equation for the instrument. I then compare the resulting models using Bayes factors. Table C.3 shows the results.

	Exclusion restriction is ...		
	violated	satisfied	satisfied with $c_{ii} = 1$
Narrative shock	log Bayes factor		
Defense spending G, Ramey (2011)	-10.6	0	-3.0
Personal income tax, Mertens and Ravn (2013)	-12.1	0	-2.8
Money shock, Romer and Romer (2004)	-5.6	0	-28.4
Money shock, Romer and Romer (2004) extended	0	-3.8	-68.2
Money shock, Romer and Romer (2004) CPI	-4.7	0	-79.5
Money shock, Kuttner (2001)	-7.3	0	-0.9
Money shock, T-Bill	-23.3	0	-191.4

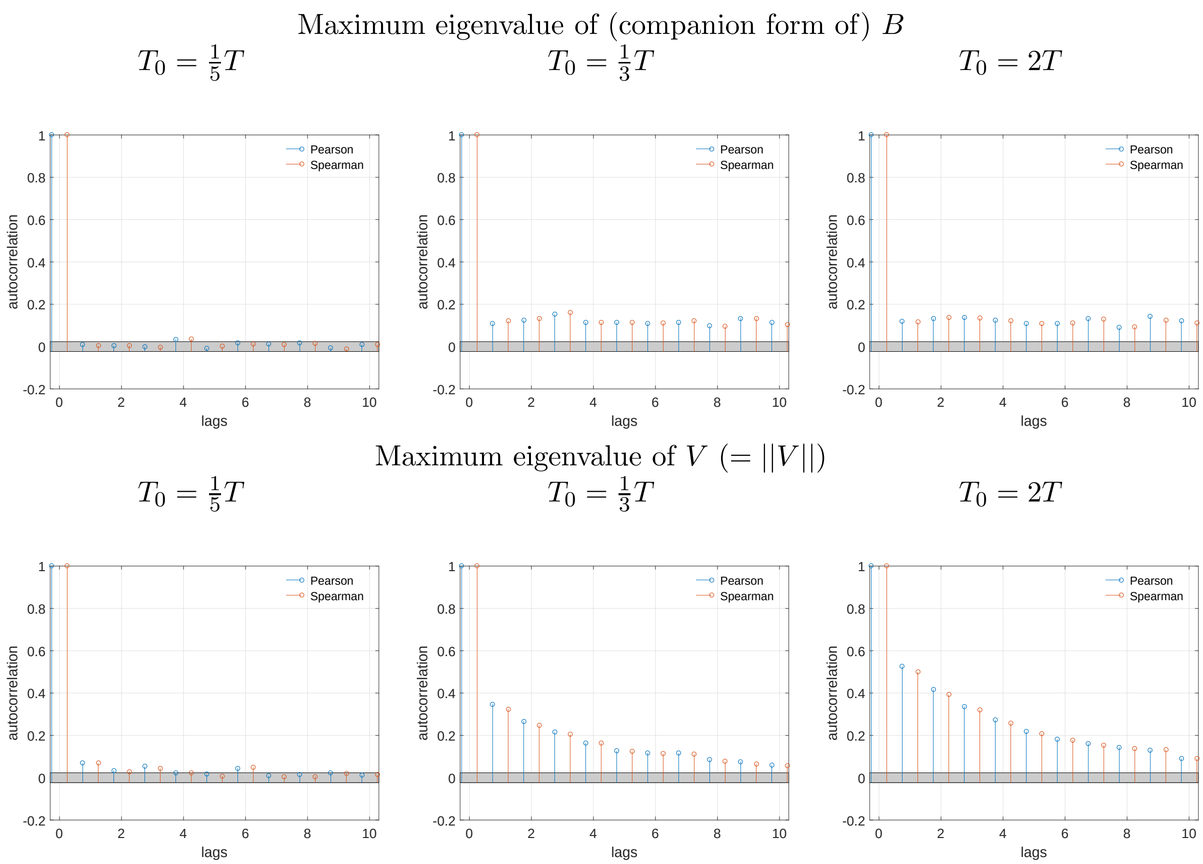
The reported Bayes factors indicate weak to very strong empirical support for models that only allow narrative shocks to affect the intended shocks themselves for the benchmark instruments reported in the upper half of the table. This is not a generic result, as the robustness in the lower half shows: With the extended Romer and Romer (2004) shock series, the most favored model relates the instruments also to shocks other than the monetary policy shock. Note: The test reports modified harmonic mean estimator of Bayes factors associated with various restrictions of equation (A.16): Narrative shocks reflect all shocks, or only the intended shock, or have, after appropriate scaling, a unit loading on the intended shock.

Table C.3: Assessing the restriction of narrative instruments to a single shock

For the narrative policy instruments used in the VAR, the data provide 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.

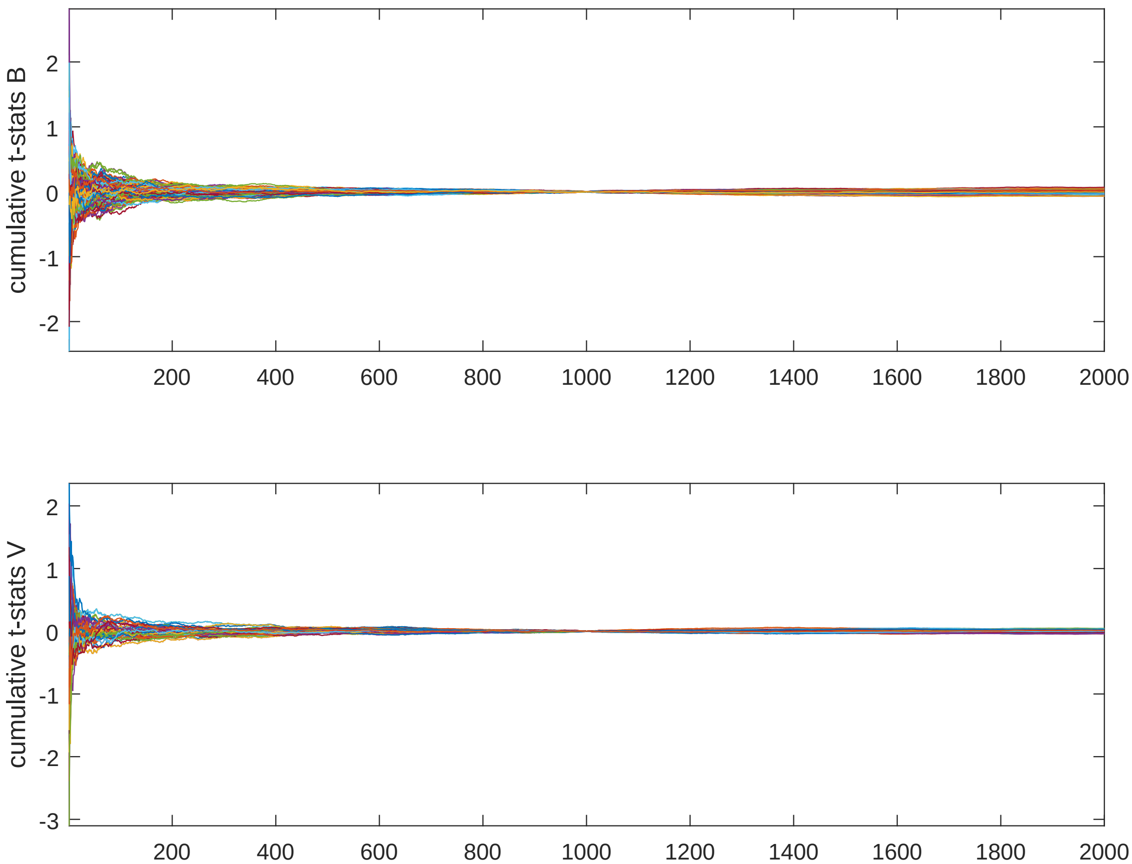
C.3 Gibbs sampler

To calibrate the Gibbs sampler, I examine the autocorrelation functions and Brooks and Gelman (1998)-type convergence statistics of all model parameters within Markov chains. See Figures C.16 and C.17 for the flat prior VAR and the DSGE-VAR, respectively. If the distributions differ visibly for different parts of the sample, 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 C.15 shows the corresponding plots. With a flat prior, I discard the first 50,000 draws and keep every 20th draw with a total accept sample of 5,000 for the DSGE-VAR and 2,000 for the flat prior VAR. This produces results consistent with convergence of the sampler (see Figures C.16 and C.17). The resulting samples are also reasonably efficient: the autocorrelation of the subsamples in Figure C.15 are reasonably small, particularly with low prior weights on the DSGE model.



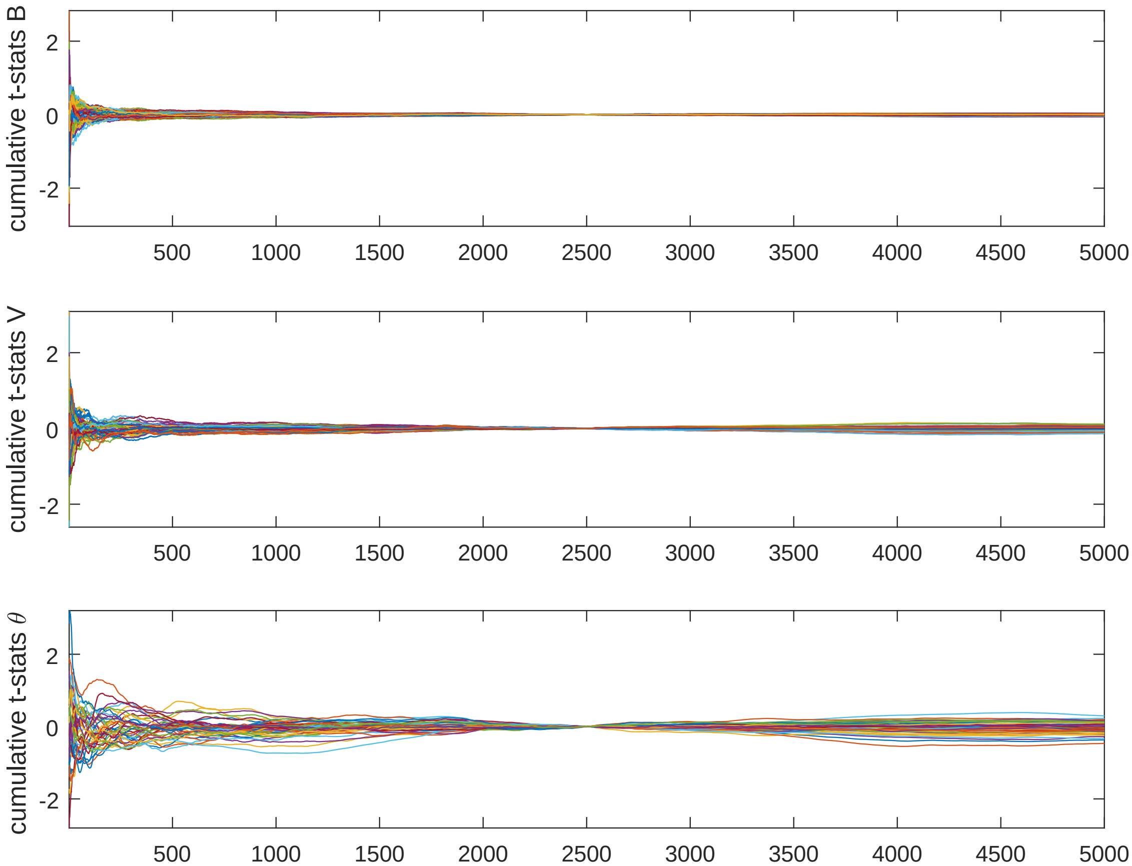
Note: Autocorrelations are reported based on both the Pearson and the Spearman correlation measure. Asymptotic classical 90% credible sets for the Pearson coefficient, computed under the assumption of zero correlation, are included around the horizontal axis. The autocorrelations are based on the thinned out sample after keeping every 20th draw. The resulting sample is reasonably efficient also with a larger prior weight on the DSGE model.

Figure C.15: Gibbs-Sampler of baseline model: autocorrelation functions of univariate summary statistics by DSGE prior weight



Shown are the (within-chain) means of the parameter estimates as the Markov chain grows. To standardize the plots, the parameter estimates are displayed minus their mean and standard deviation in the first half of the chain: For example, for element i of θ, the plot shows $\frac{t^{-1}\sum_{s=1}^{t}\theta_s(i)-\lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}\theta_s(i)}{\left(\lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}\left(\theta_s(i)-\lfloor T/2\rfloor^{-1}\sum_{u=1}^{\lfloor T/2\rfloor}\theta_u(i)\right)^2\right)^{1/2}}$ as a function of t. Brooks and Gelman (1998) argue that these means should have converged for a satisfying posterior simulation. The results above indicate that the convergence is very good for both the elements of the VAR coefficient matrix B and the covariance matrix V.

Figure C.16: Brooks and Gelman (1998) type convergence diagnostic for the flat-prior narrative VAR



Shown are the (within-chain) means of the parameter estimates as the Markov chain grows. To standardize the plots, the parameter estimates are displayed minus their mean and standard deviation in the first half of the chain: For example, for element i of θ, the plot shows $\frac{t^{-1}\sum_{s=1}^{t}\theta_s(i)-\lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}\theta_s(i)}{\left(\lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}\left(\theta_s(i)-\lfloor T/2\rfloor^{-1}\sum_{u=1}^{\lfloor T/2\rfloor}\theta_u(i)\right)^2\right)^{1/2}}$ as a function of t. Brooks and Gelman (1998) argue that these means should have converged for a satisfying posterior simulation. The results above indicate that the convergence is best for the elements of the VAR coefficients B and almost as good for the elements of the covariance matrix V. Some structural parameter draws seem to only settle down after about 4,000 draws.

Figure C.17: Brooks and Gelman (1998) type convergence diagnostic for DSGE-VAR with $T_0 = 1$