A Narrative Approach to a Fiscal DSGE Model
Online Appendix
This version: December 12, 2019
First version: February 2014
Thorsten Drautzburg*
Given that

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

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

\[ \Sigma = \begin{bmatrix} A_{11}A_{11}' + A_{12}A_{12}' & A_{11}A_{11}' + A_{12}A_{22}' \\ \kappa A_{11}A_{11}' + A_{22}A_{12}' & \kappa A_{11}A_{11}' + A_{22}A_{22}' \end{bmatrix} \]

(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 \( A_{11} \):

\[ u_t = A\varepsilon_t = [A^{[1]} A^{[2]}] \varepsilon_t = A^{[1]}\varepsilon_t^{[1]} + A^{[2]}\varepsilon_t^{[2]} = \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} A_{11}e_t^{[1]} + \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} e_t^{[2]} \]

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

\[ \text{Var}[u_t|\varepsilon_t^{[1]}] = A^{[2]}(A^{[2]})' = \begin{bmatrix} A_{12}A_{12}' & A_{12}A_{22}' \\ A_{22}A_{12}' & A_{22}A_{22}' \end{bmatrix} \]

\[ \text{Var}[u_t|\varepsilon_t^{[2]}] = A^{[1]}(A^{[1]})' = \begin{bmatrix} A_{11}A_{11}' & A_{11}A_{11}' \\ \kappa A_{11}A_{11}' & \kappa A_{11}A_{11}' \end{bmatrix} \]

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

Note that:

\[ u_t^{res} = u_t - \mathbb{E}[u_t|\varepsilon_t^{[1]}] \perp \mathbb{E}[u_t|\varepsilon_t^{[1]}] = \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} A_{11}e_t^{[1]} \]

Any vector in the nullspace of \( [I_{m_z} \kappa]' \) satisfies the orthogonality condition.

Note that \( \{ [I_{m_z} \kappa]', [-I_{m-m_z}] \} \) is an orthogonal basis for \( \mathbb{R}^m \).

Define

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

(A.3)

Note that \( Z^{[2]} \) spans the nullspace of \( A^{[1]}' \). Hence, \( (Z^{[2]})'v_t \) projects \( v_t \) onto the nullspace of the instrument-identified shocks \( \varepsilon_t^{[1]} \).

\[ (Z^{[2]})'v_t = (Z^{[2]})'A\varepsilon_t = (Z^{[2]})' [A^{[1]} A^{[2]}] \varepsilon_t = \begin{bmatrix} \Sigma_{12}\Sigma_{12}' & \Sigma_{12}\Sigma_{22}' \\ \Sigma_{22}\Sigma_{12}' & \Sigma_{22}\Sigma_{22}' \end{bmatrix} \varepsilon_t = [0 \ (Z^{[2]})'A^{[2]}] \varepsilon_t = 0 \times \varepsilon_t^{[1]} + (Z^{[2]})'A^{[2]}\varepsilon_t^{[2]} \perp \varepsilon_t^{[1]} \]

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

$$E[v_t|\varepsilon_t^{[2]}] = \text{Cov}[v_t, (Z^{[2]})'v_t] \text{Var}[(Z^{[2]})'v_t]^{-1}(Z^{[2]})'v_t,$$

$$v_t - E[v_t|\varepsilon_t^{[2]}] = (I - \text{Cov}[v_t, (Z^{[2]})'v_t] \text{Var}[(Z^{[2]})'v_t]^{-1}(Z^{[2]})'v_t,$$

$$\text{Cov}[v_t, (Z^{[2]})'v_t] = \Sigma Z^{[2]} = \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix}.$$

$$\text{Var}[v_t|\varepsilon_t^{[2]}] = \text{E}[(I - \text{Cov}[v_t, (Z^{[2]})'v_t] \text{Var}[(Z^{[2]})'v_t]^{-1}(Z^{[2]})'v_t]v_t]$$

$$= \text{E}[v_t v_t'] - \text{Cov}[v_t, (Z^{[2]})'v_t] \text{Var}[(Z^{[2]})'v_t]^{-1}\text{E}[(Z^{[2]})'v_t]v_t'$$

$$= \Sigma - \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \left[ \kappa - I_{m-m_z} \right] \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix}^{-1} \left[ \kappa - I_{m-m_z} \right] \Sigma$$

$$= \begin{bmatrix} A_{11} A_{11}' & A_{11} A_{11}' \\ \kappa A_{11} A_{11}' \end{bmatrix} \text{A.4}$$

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

$$\text{Var}[v_t|\varepsilon_t^{[1]}] = \Sigma - \text{Var}[v_t|\varepsilon_t^{[2]}]$$

$$= \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix} \left[ \kappa - I_{m-m_z} \right] \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_z} \end{bmatrix}^{-1} \left[ \kappa - I_{m-m_z} \right]' \Sigma \text{A.5}$$

In general, $A_{11}$ 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 $A_{11} A_{11}'$. Given $A_{11}$, the impact response to a unit shock is given by:

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

### 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 (3.4). Recall the definition of forecast errors $v_t$ in terms of structural shocks $\varepsilon_t$:

$$v_t = A\varepsilon_t \equiv \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \Leftrightarrow \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \text{A.6}$$

$$v_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

Note that:

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

$$= \begin{bmatrix} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} & -A_{11}^{-1} A_{12} (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} \\ -A_{22}^{-1} A_{21} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} & (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} \end{bmatrix}$$
Note that
\[
(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1}((A_{11} - A_{12}A_{22}^{-1}A_{21})A_{11}^{-1})^{-1} = A_{11}^{-1}(I - A_{12}A_{22}^{-1}A_{21}A_{11}^{-1})^{-1}
\]
and define:
\[
S_1 \equiv (I - A_{12}A_{22}^{-1}A_{21}A_{11}^{-1})A_{11} \quad S_2 \equiv (I - A_{21}A_{11}^{-1}A_{12}A_{22}^{-1})A_{22}
\]
so that
\[
(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = S_1^{-1} \quad (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} = S_2^{-1}
\]

Using these equalities gives the first equality in what follows, whereas the second equality is straightforward algebra:
\[
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}^{-1} v_t = \begin{bmatrix}
S_1^{-1} & -S_1^{-1}A_{11}^{-1}A_{12} \\
-S_2^{-1}A_{22}^{-1}A_{21} & S_2^{-1}
\end{bmatrix} v_t = \begin{bmatrix}
I & -A_{12}A_{22}^{-1} \\
0 & I
\end{bmatrix} v_t = \begin{bmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{bmatrix}
\]
and equivalently:
\[
\begin{bmatrix}
I & -\eta \\
-\kappa & I
\end{bmatrix} v_t = \begin{bmatrix}
S_1 & 0 \\
0 & S_2
\end{bmatrix} \begin{bmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{bmatrix} \tag{A.8}
\]
defining \(\eta \equiv A_{12}A_{22}^{-1}\) and \(\kappa \equiv A_{21}A_{11}^{-1}\). Equation (3.4) follows immediately.

Lemma 1 (Mertens and Ravn (2013)). Let \(\Sigma = AA'\) and \(\Gamma = [G \ 0] A'\), where \(G\) is an \(m_z \times m_z\) invertible matrix and \(A\) is of full rank. Then \(A^{[1]}\) is identified up to a factorization of \(S_1S'_1\) 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 \(\eta, \kappa\), (A.8) implies (3.5), which I reproduce here for convenience:
\[
A^{[1]} = \begin{bmatrix}
A_{11} \\
A_{21}
\end{bmatrix} = \begin{bmatrix}
(I - \eta\kappa)^{-1} \\
(I - \kappa\eta)^{-1}\kappa
\end{bmatrix} \text{chol}(S_1S'_1). \tag{3.5}
\]
If \(\Sigma\) and \(\Gamma\) pin down \(\eta, \kappa\) uniquely, \(A^{[1]}\) is uniquely identified except for a factorization of \(S_1S'_1\).

To show that \(\Sigma\) and \(\Gamma\) pin down \(\eta, \kappa\) uniquely, consider \(\kappa\) first. Since \(\Gamma = [G \ 0] A\) and \(G\) is an \(m_z \times m_z\) invertible matrix, it follows that Assumption 1 holds. It then follows from (3.2) that \(\kappa = A_{21}A_{11}^{-1} = \Gamma^2\Gamma_1^{-1}\).

To compute \(\eta\), more algebra is needed. Partition \(\Sigma = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{12}' & \Sigma_{22}
\end{bmatrix}\), where \(\Sigma_{11}\) is \(m_z \times m_z\), \(\Sigma_{12}\) is \(m_z \times (m - m_z)\) and \(\Sigma_{22}\) is \((m - m_z) \times (m - m_z)\). Define
\[
A_{22}A_{22}' = \Sigma_{22} - \kappa A_{11}A_{11}'\kappa' = \Sigma_{22} - \kappa(\Sigma_{11} - A_{12}A_{12}')\kappa',
\]
using (A.2) twice. Using the upper left element of (A.5), it follows that
\[
A_{12}A_{12}' = (\Sigma_{12}' - \kappa\Sigma_{11})'(ZZ)'^{-1}(\Sigma_{12}' - \kappa\Sigma_{11})
\]
Thus, $\eta$ is then defined as:

\[
\eta \equiv A_{12}A_{22}^{-1} = A_{12}A_{22}(A_{22}A_{22}')^{-1} = (\Sigma_{12} - \kappa A_{11}A_{11}') (A_{22}A_{22}')^{-1} \\
= (\Sigma_{12} - \kappa\Sigma_{11}' + \kappa A_{12}A_{12}') (A_{22}A_{22}')^{-1}.
\]

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

\[
\eta = A_{12}A_{22}^{-1} = A_{12}A_{22}(A_{22}A_{22}')^{-1} = (\Sigma_{12} - \kappa A_{11}A_{11}') (A_{22}A_{22}')^{-1}.
\]

Thus, $\eta$ and $\kappa$ are uniquely identified given $\Sigma, \Gamma$.  

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

**Proposition 1.** Assume $\Sigma = AA' = A^*(A^*)'$ and order the policy variables such that the $m = m_z$ or $m = m_z - 1$ observable Taylor rules are ordered first and $\Gamma = [G, 0]A^*$. Then $A^{[1]}$ defined in (3.5) satisfies $A^{[1]} = A^*[I_{m_z}, 0_{(m-m_z)\times (m-m_z)}]'$ up to a normalization of signs on the diagonal if

(a) $m_z$ instruments jointly identify shocks to $m = m_z$ observable Taylor rules w.r.t. the economy (2.2), or

(b) $m_z$ instruments jointly identify shocks to $m = m_z - 1$ observable Taylor rules w.r.t. the economy (2.2) and $\eta_{m_z} = 0, p = 1, \ldots, m_p$.

**Proof.** Given Lemma 3, $A^{[1]}$ 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 $A^{[1]}$ is identified uniquely and, hence, equal to $A^*[I_{m_z}, 0_{(m-m_z)\times (m-m_z)}]'$.

To proceed, stack the $m_p$ policy rules:

\[
y_t^p = \sum_{i=m_p+1}^{m} \begin{bmatrix} \eta_1,i & 0 & \ldots & 0 \\ 0 & \eta_2,i & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \eta_{m_p,i} \end{bmatrix} y_{i,t} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{m_p} \end{bmatrix} x_{t-1} + \begin{bmatrix} \Sigma_{11} & 0 & \ldots & 0 \\ \Sigma_{21} & \Sigma_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ n_{p,1} & \sigma_{n_p,2} & \ldots & \sigma_{n_p,n_p} \end{bmatrix} \epsilon_t, \\
= \sum_{i=m_p+1}^{m} D_iy_{i,t} + A x_{t-1} + [D_0 \ 0] \epsilon_t,
\]

where $m - m_p = \bar{n} \equiv \max_p n_p$. Define $e_i$ as the selection vector of zeros except for a one at its $i$th position and denote the $i$th row of matrix $A$ by $A_i = (e_i'A')$ 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 observation equation (2.2a) can be written as:

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

where $0$ 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 \([A^* I_m] \xrightarrow{E} [B | C] = [EA^* | EI_m]\). Then:

\[
[(A^*)|I_m] = \begin{bmatrix}
[D_0 \ 0] + \sum_{i=m_p+1}^{m} D_i 1(A^*_i)' & (A^*_{m_p+1})' \\
\vdots & \vdots \\
(A^*_m)' & 0' \\
\end{bmatrix}
\]

Thus, \((A^*)^{-1}\) is lower triangular. Since the lower Cholesky decomposition is unique, a Cholesky

Thus, \((A^*)^{-1}\) is lower triangular. Since the lower Cholesky decomposition is unique, a Cholesky

Thus, \((A^*)^{-1}\) 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 \( A^{[l]} \) follows from Lemma 3.

A.3 VAR priors and posteriors

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

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

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

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

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

The likelihood can then also be written as:

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

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 \( \beta: \beta | V^{-1} \sim \mathcal{N}(\tilde{\beta}_{SUR}, (\tilde{X}_{SUR}' \tilde{X}_{SUR}^{-1})^{-1}) \equiv \mathcal{N}(\tilde{\beta}_{SUR}, (X_{SUR}' (V^{-1} \otimes I_T) X_{SUR})^{-1}) \). 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:

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

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

(A.11)

which is $V^{-1}|\beta \sim W_{ma}((SSR(\beta))^{-1}, T)$, with

$$
SSR(\beta) \equiv U(\beta)'U(\beta) = [Y - X_y B_y(\beta), Z - X_z B_z(\beta)]' [Y - X_y B_y(\beta), Z - X_z B_z(\beta)]
$$

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

A.4 Invertibility

A necessary condition for the VAR and DSGE models to agree on the structural shocks is that both models 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 (2.2) matches those from the VAR (2.1):

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

Under this condition, the forecast errors of a VAR with sufficiently many lags and the DSGE model coincide, as summarized in the following lemma:

**Lemma 2.** (Fernandez-Villaverde et al., 2007, p. 1022) Let $y_t$ be generated by the DSGE economy (2.2). Under Assumption 2, the variance-covariance matrix of the one-step-ahead prediction error in the Wold representation of $y_t$ is given by $\Sigma^* = (A^*)(A^*)'$.

In my application, 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 $y_t$. With an equal number of AR(1) shock processes as observables, I found two intuitive cases in my exploratory analysis that violate Assumption 2 for most of the parameter space: First, a model with capital that exclude investment and capital as observables. This is similar to Chari et al. (2005) who point to the challenge of recovering impulse responses in VAR models in economies with capital. Second, a model with news shocks and without observed expectations. For the estimated models, however, I show in Appendix B.2 that a VAR(4) approximation captures the underlying DSGE model dynamics well.

A.5 DSGE-SVAR 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 $\mathcal{N}(\beta_0, V_0^{-1})$ distribution, whereas the observations that generate the prior for the covariance matrix are generated from a $\mathcal{N}(0, V^{-1})$ distribution.

---

1Intuitively, $x_t$ can be expressed as a square-summable linear combination in terms of $y$. Hence, $\text{Var}[x_t|y^t] = 0$ and the Wold representation of $y_t$ is given by: $y_t = B^* \sum_{j=0}^{\infty} (C^* - D^*(A^*)^{-1}B^*)(A^*)^{-1}y_{t-1-j} + A^*\epsilon_t$. The one-step-ahead prediction error is, therefore, $y_t - E[y_t|y^{t-1}] = A^*\epsilon_t$ with variance $(A^*)(A^*)'$.

2I also verify this condition for each draw of the DSGE model parameters in my empirical application.
Specify:

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

Note that this is not equal to

$$\beta | \mathbf{V}^{-1} \sim \mathcal{N}(\bar{\beta}_0, \mathbf{N}_0(\mathbf{V}^{-1}))$$

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

The prior for $\mathbf{V}^{-1}$ is Wishart independent of $\beta$.

$$\mathbf{V}^{-1} \sim \mathcal{W}_{m+m_2}(\mathbf{V}_0 T_0, T_0)$$

Note that because the prior for $\beta$ is independent of $\mathbf{V}^{-1}$, the prior is conditionally conjugate with the likelihood function. Otherwise, the presence of $|\mathbf{N}_0(\mathbf{V}^{-1})|$ terms would undo the conjugacy.

The prior is therefore:

$$
\pi(\beta, \mathbf{V}^{-1}|\theta) = (2\pi)^{-n/2} |\bar{N}_0(\theta)|^{-1/2} e^{-\frac{1}{2} (\beta - \bar{\beta}_0(\theta))' \bar{N}_0(\theta)(\beta - \bar{\beta}_0(\theta))}
\times 2^{-T_0(m + m_2)/2} |\mathbf{V}_0(\theta) T_0|^{-T_0/2} e^{-\frac{1}{2} \text{tr}(\mathbf{V}^{-1} \mathbf{V}_0(\theta) T_0)}
= (2\pi)^{-n/2} |\bar{N}_0(\theta)|^{-1/2} e^{-\frac{1}{2} (\beta - \bar{\beta}_0(\theta))' \mathbf{N}_0(\theta)(\beta - \bar{\beta}_0(\theta))}
\times 2^{-T_0(m + m_2)/2} |\mathbf{S}_0(\theta)|^{-T_0/2} e^{-\frac{1}{2} \text{tr}(\mathbf{V}^{-1} \mathbf{S}_0(\theta))}
$$

The joint density is given by:

$$p(\mathbf{Y}, \mathbf{Z}, \beta, \mathbf{V}^{-1}, \theta) = p(\mathbf{Y}, \mathbf{Z}|\beta, \mathbf{V}^{-1}) p(\beta, \mathbf{V}^{-1}|\theta) p(\theta),
= p(\mathbf{Y}, \mathbf{Z}|\beta, \mathbf{V}^{-1}) p(\beta|\theta) p(\mathbf{V}^{-1}|\theta) p(\theta) \quad (A.12a)$$

$$p(\mathbf{Y}, \mathbf{Z}|\beta, \mathbf{V}^{-1}) = (2\pi)^{-T/2} \mathbf{V}^{-1} \mathbf{V}^{-1/2} \text{vec}(\mathbf{Y}, \mathbf{Z})' \mathbf{X}_{SUR}(\beta) (\mathbf{V}^{-1} \otimes \mathbf{I}_T) (\text{vec}(\mathbf{Y}, \mathbf{Z}) - \mathbf{X}_{SUR}(\beta)) \quad (A.12b)$$

$$p(\beta|\theta) = (2\pi)^{-n/2} \left| \lambda_B \mathbf{X}'_{0,SUR}(\theta) (\mathbf{V}_0(\theta)^{-1} \otimes \mathbf{I}_{m(mp + k)}) \mathbf{X}_{0,SUR} \right|^{1/2}
\times e^{-\frac{1}{2} \text{tr}(\mathbf{A}^{-1} \mathbf{V}_0(\theta))}
$$

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

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

$$p(\mathbf{V}^{-1}|\theta) = e^{-\frac{1}{2} \text{tr}(\mathbf{V}^{-1} \mathbf{V}_0(\theta))} \frac{|T_0^V \mathbf{V}_0(\theta)|^{T_0^V/2}}{2^{T_0^V(m + m_2)/2} \Gamma_{m+m_2} \left(\frac{T_0^V}{2}\right)}
\mathbf{V}^{-1/2} \mathbf{V}_0(\theta) \mathbf{V}_0(\theta)', \quad (A.12d)$$

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

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

**Implementing the DSGE-SVAR prior.** The following dummy observations and likelihood implement the prior that $\beta \sim \mathcal{N}(\bar{\beta}_0, \mathbf{N}_{XX}(\mathbf{V}_0))$ and $\mathbf{V}^{-1} \sim \mathcal{W}(\mathbf{V}_0 T_0^V, T_0^V)$:

$$\text{vec}([\mathbf{Y}_0^B, \mathbf{Z}_0^B]) = \bar{\mathbf{X}}_{0,SUR}(\theta) \bar{\beta}_0(\theta) + \mathbf{0}, \quad (A.13a)$$

$$\text{vec}([\mathbf{Y}_0^B, \mathbf{Z}_0^B]) \sim \mathcal{N}(\bar{\mathbf{X}}_{0,SUR}(\theta) \bar{\beta}_0(\theta), \mathbf{V}_0(\theta) \otimes \mathbf{I}_{T_0^B}) \quad (A.13b)$$
\[ [\mathbf{Y}_0^V, \mathbf{Z}_0^V] = \mathbf{0} \times \beta + \mathbf{V}_0(\theta) \otimes \mathbf{I}_{T_0^V}, \]  
(A.13c)

\[
\text{vec}([\mathbf{Y}_0^V, \mathbf{Z}_0^V]) \sim \mathcal{N}(\mathbf{0}, \mathbf{V} \otimes \mathbf{I}_{T_0^V})
\]  
(A.13d)

where \( \mathbf{X}_{0,SUR} \) is the Cholesky factor of the following matrix:

\[
\mathbf{X}_{0,SUR}(\theta)\mathbf{X}_{0,SUR}(\theta) = E^{DSGE} [\mathbf{X}_{SUR}(\mathbf{V}(\theta))^{-1} \otimes \mathbf{I}_{p(m+m_z)}] \mathbf{X}_{SUR}(\theta).
\]

The prior and data density are given by:

\[
\theta \sim \pi(\theta) \quad \text{(A.14a)}
\]

\[
\pi(\mathbf{B}, \mathbf{V}^{-1} | \theta) \propto |\mathbf{V}^{-1}|^{-n_y/2} f(\mathbf{B}, \mathbf{V}^{-1} | \mathbf{Y}_0(\theta), \mathbf{Z}_0(\theta)) = |\mathbf{V}^{-1}|^{-n_y/2} f(\mathbf{Y}_0(\theta), \mathbf{Z}_0(\theta) | \mathbf{B}, \mathbf{V}^{-1}) \quad \text{(A.14b)}
\]

\[
\tilde{f}(\mathbf{Y}, \mathbf{Z}| \mathbf{B}, \mathbf{V}^{-1}, \theta) = f(\mathbf{Y}, \mathbf{Z} | \mathbf{B}, \mathbf{V}^{-1}) \quad \text{(A.14c)}
\]

**Computations.** To implement Algorithm 3, I use a random-blocking Metropolis-Hastings step with random walk proposal density with t-distributed increments, with 15 degrees of freedom as in Chib and Ramamurthy (2010). 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.

**A.6 Results on the marginal data density in \( T_0^V \)**

**A.6.1 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.

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

\[
p(y|T_0^V) = \int_0^\infty \left( \prod_{t=1}^T f(y_t|V) \right) \pi(V|T_0^V) dV^{-1}
\]

\[
= \int_0^\infty (2\pi)^{-mT/2} |V|^{-T/2} e^{-\frac{1}{2} tr(V^{-1}\tilde{V})} \frac{T_0^V}{2} tr(V^{-1}V_0) 2^{-mT_0^V/2} |V_0 V_0^T|^{m}/2 \Gamma_m(T_0^V/2)^{-1} |V|^{-(T_0-m-1)/2} dV^{-1}
\]

\[
= \pi^{-mT/2} \frac{|V_0 V_0^T|^{m}/2 \Gamma_m((T + T_0^V)/2)}{|V_0 V_0^T + TV|^{(T + T_0^V)/2} \Gamma_m(T_0^V/2)} \times \int_0^\infty e^{-\frac{1}{2} tr(V^{-1}(TV + T_0^V V_0))} 2^{-m(T + T_0^V)/2} |V_0 V_0^T + TV|^{(T + T_0^V)/2} \Gamma_m((T + T_0^V)/2)^{-1} |V|^{-(T + T_0 - 2)/2} dV^{-1}
\]

\[
= \pi^{-mT/2} \frac{|V_0 V_0^T|^{m}/2 \Gamma_m((T + T_0^V)/2)}{|V_0 V_0^T + TV|^{(T + T_0^V)/2} \Gamma_m(T_0^V/2)}
\]
\[
\frac{dT_0^V}{d} = \frac{1}{2} \ln \left( \frac{T_0^V V_0}{T_0^V + T} \right) + \frac{1}{2} \left( 1 - \frac{V_0}{V} \right) \frac{T}{T_0^V V_0 + T} + \frac{1}{2} \psi \left( \frac{T + T_0^V}{2} \right) - \frac{1}{2} \psi \left( \frac{T_0^V}{2} \right), \quad (A.16)
\]

where \( \psi \) is the digamma function, the derivative of the log Gamma function. Part (a) of the following Lemma establishes that for \( \frac{V_0}{V} \) 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 \( V_0 = \hat{V} \) 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 \( \frac{V_0}{V} \) 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 3.** Let \( T = 2n, n \in \mathbb{N}_+ \) and \( T_0^V > 0 \).

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

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

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

**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

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

The slope \( (A.16) \) can therefore be written as:

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

Note that for \( x > 0 \):

\[
\frac{1}{2} \ln \left( \frac{x}{2 + x} \right) + \frac{1}{x} > 0.
\]
A.6.2 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 $\bar{B}_0$, $\bar{B}_0^B$, 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 $T_0^B = T_0^V \to \infty$.

A.7 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.
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

**Likelihood given DSGE parameters.** The basic insight from Chib (1995) is that:

\[
\pi(y, z|\theta) = \frac{p(y, z|V^{-1}, B)\pi(V^{-1}, B|\theta)}{\pi(\hat{z}, V^{-1}, B|y, z, \theta)} = \frac{p(y, z|V^{-1}, B)\pi(V^{-1}, B|\theta)}{\pi(B_s|y, z, \theta)\pi(V_s^{-1}|B_s, y, z, \theta)},
\]

(A.17)

for any \( V^{-1}, B \). For numerical purposes, however, it is advisable to evaluate (A.17) at a high density point. In what follows I denote this point by \((\hat{z}_s, B_s, V_s^{-1})\). I choose \( B_s \) as the posterior mean. I first compute:

\[
\pi(B_s|y, z, \theta) = M^{-1} \sum_{m=1}^{M} \pi(B_s|y, z, (V_s^{-1})^{(m)}, \hat{z}^{(m)}, \theta),
\]

using draws \( \{(V_s^{-1})^{(m)}, \hat{z}^{(m)}\} \) from the original Gibbs sampler. The second component is computed as:

\[
\pi(V_s^{-1}|y, z, \theta) = M^{-1} \sum_{m=1}^{M} \pi(V_s^{-1}|y, z, B_s, \hat{z}^{(m)}, \theta),
\]

where \( (V_s^{-1})^{(m)}, \hat{z}^{(m)} \) are draws from a simpler new run of the Gibbs sampler that conditions on \( B_s \).

To compute the likelihood \( p(y, z|B_s, V_s^{-1}) \), I draw a third sequence of \( \hat{z}^{(m)} \) conditional on both \( B_s, V_s^{-1} \) and I compute \( p(y, z|B_s, V_s^{-1}) = M^{-1} \sum_{m=1}^{M} p(y, z, \hat{z}^{(m)}|B_s, V_s^{-1}) \).

**Likelihood over DSGE parameters.** Geweke (1999) shows that to find the integrating constant of a Kernel \( k(\psi) \) we may use that \( p(\hat{g}) \) is the integrating constant of the posterior kernel \( k(\psi) = \)
Let \( g(\psi) = \frac{f(\psi)}{k(\psi)} \). Then:

\[
\mathbb{E}[g(\psi)] = \int f(\psi) k(\psi) p(\psi | y) d\psi = p(y)^{-1} \int f(\psi) k(\psi) d\psi = p(y)^{-1} \int f(\psi) d\psi = p(y)^{-1}
\]

for any density \( f(\psi) \). Geweke (1999) proposes to use a truncated normal density function with the posterior mean and covariance of \( \psi \). Denote this truncated density by \( f_1(\psi) \) and its estimate based on the sample posterior distribution with sample size \( M \) by \( f_{\alpha, M}(\psi_m) \). Then:

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

where \( \psi_m \) are draws from the posterior.

Here, \( \psi = (\theta, B, V^{-1}) \) -- or strictly \( (\theta, B, \text{vech}(V^{-1})) \). 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, which I do using the Chib (1995) algorithm previously described.

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

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

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

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

\[
\Leftrightarrow k(\theta) = p(y, z, \theta) \pi(\theta)
\]

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 \( \int_{\Theta} f(\theta) p(y, z, \theta) d\theta \) 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.8 DSGE model equations

A.8.1 Households

The law of motion for capital:

\[
\dot{k}_t^p = (1 - \frac{x}{k^p})\dot{k}_{t-1}^p + \frac{x}{k^p}(\dot{x}_t + \dot{q}_{t+s})
\]  

(A.18)

Household wage setting:

\[
\dot{w}_t = \frac{\dot{w}_{t-1}}{1 + \beta \gamma} + \frac{\beta \gamma E_t[\hat{w}_{t+1}]}{1 + \beta \gamma} + \frac{(1 - \beta \zeta_w)(1 - \zeta_w) - A_w}{(1 + \beta \gamma)\zeta_w} \left( \frac{\hat{c}_t - (h/\gamma)\hat{c}_{t-1}}{1 - h/\gamma} + \nu \hat{t}_t - \dot{w}_t + \frac{d\tau^n_t}{1 - \tau^n} + \frac{d\tau^c_t}{1 + \tau^c} \right)
\]

\[
- \frac{1 + \beta \mu w_t}{1 + \beta \gamma} \hat{\pi}_t + \frac{\nu w_t}{1 + \beta \gamma} \hat{\pi}_{t-1} + \frac{\beta \gamma}{1 + \beta \gamma} E_t[\hat{\pi}_{t+1}] + \frac{\dot{\varepsilon}_t w}{1 + \beta \gamma}
\]  

(A.19)

Household consumption Euler equation:

\[
E_t[\hat{\xi}_{t+1} - \hat{\xi}_t] + E_t[d\tau^c_{t+1} - d\tau^c_t] =
\]

\[
= \frac{1}{1 - h/\gamma} E_t((\sigma - 1) \frac{1}{1 + \lambda w} \frac{1 - \tau^n}{1 + \tau^c} \frac{\hat{w}_t}{c} \hat{\pi}_{t+1} - \hat{n}_t - \sigma \left[ \hat{c}_{t+1} - \left(1 + \frac{h}{\gamma}\right) c_t + \frac{h}{\gamma} \hat{c}_{t+1} \right]),
\]  

(A.20)

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

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

(A.21)

\[
\dot{Q}_t = -\dot{q}_t^b - (\dot{R}_t - E_t[\hat{\pi}_{t+1}]) + \frac{1}{\bar{r}^k(1 - \tau^k) + \delta \tau^k + 1 - \delta} \times
\]

\[
\left[ (\bar{r}^k(1 - \tau^k) + \delta \tau^k)\dot{q}_t^k - (\bar{r}^k - \delta)\tau_{t+1}^k + \bar{r}^k(1 - \tau^k)E_t(\dot{q}_t^k) + (1 - \delta)E_t(\dot{Q}_{t+1}) \right],
\]  

(A.22)

\[
\dot{x}_t = \frac{1}{1 + \beta \gamma} \left[ \dot{x}_{t-1} + \beta \gamma E_t(\dot{x}_{t+1}) + \frac{1}{\gamma^2 S^u(\gamma)} \left[ \dot{Q}_t + \dot{q}_t^b \right] \right],
\]  

(A.24)

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

(A.25)

A.8.2 Production side and price setting

The linearized aggregate production function is:

\[
\hat{y}_t = \frac{\bar{y} + \Phi}{\bar{y}} \left( \hat{c}_t + \zeta \hat{k}_{t-1}^q + \alpha (1 - \zeta) \hat{k}_t + (1 - \alpha) (1 - \zeta) \hat{n}_t \right),
\]  

(A.26)

where \( \Phi \) are fixed costs. Fixed costs, in steady state, equal the profits made by intermediate producers.

The capital-labor ratio:

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

(A.27)
Price setting:
\[ \hat{\pi}_t = \frac{\iota_p}{1 + t_p \beta \gamma} \hat{\pi}_{t-1} + \frac{1 - \zeta_p}{1 + t_p \beta \gamma} \hat{\pi}, \]
\[ + \frac{\zeta_p}{1 + t_p \beta \gamma} \bar{A}_p (\hat{m}_t + \epsilon_t^{Xp}) + \frac{\beta \gamma}{1 + t_p \beta \gamma} \bar{E}_t \hat{\pi}_{t+1}. \] (A.28)

Marginal costs with a cost-channel:
\[ \hat{m}_t = \alpha \hat{x}_t^k + (1 - \alpha)(\hat{w}_t + \hat{R}_t). \] (A.29)

A.8.3 Market clearing
Goods market clearing requires:
\[ \hat{y}_t = \bar{c} \hat{c}_t + \bar{x} \hat{x}_t + \bar{g} \hat{g}_t + \frac{\bar{\tau}_y}{\bar{y}} \hat{u}_t. \] (A.30)

A.8.4 Observation equations
For the estimation under full information, I need to specify observation equations. The observation equations are given by (3.1c) as well as the following seven observation equations from Smets and Wouters (2007) and three additional equations (A.32) on fiscal variables:
\[ \Delta \ln g^o_t = g_t - g_{t+1} + (\gamma_g - 1), \] (A.31a)
\[ \Delta \ln x^o_t = x_t - x_{t+1} + (\gamma_x - 1), \] (A.31b)
\[ \Delta \ln w^o_t = w_t - w_{t+1} + (\gamma_w - 1), \] (A.31c)
\[ \Delta \ln c^o_t = c_t - c_{t+1} + (\gamma - 1), \] (A.31d)
\[ \hat{\pi}^o_t = \hat{\pi}_t + \pi, \] (A.31e)
\[ \hat{n}^o_t = \hat{n}_t + \bar{n}, \] (A.31f)
\[ \hat{R}^o_t = \hat{R}_t + (\beta^{-1} - 1), \] (A.31g)

By allowing for different trends in the non-stationary observables, I treat the data symmetrically in the VAR and the DSGE model.
I use the deviation of debt to GDP and revenue to GDP, detrended prior to the estimation, as observables:
\[ b^o_t = \frac{\bar{b}}{\bar{y}} (\hat{b} - \bar{y}) + \bar{b}^o \] (A.32a)
\[ rev^{n,o}_t = \frac{\bar{n}}{\bar{y}} (\hat{w} + \hat{n}_t - \bar{y}) + \bar{rev}^{n,o} \] (A.32b)
\[ rev^{k,o}_t = \frac{\bar{k}}{\bar{y}} (\hat{r} - \delta) (\frac{\bar{r}}{\bar{r} - \hat{r}} + \hat{r}_t - \hat{k}_t - \bar{y}) + \bar{rev}^{k,o} \] (A.32c)

B Data and additional results
B.1 Data construction
NIPA and Flow of Funds variables. 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:

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

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

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

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

\[
r_t = \left\{ \begin{array}{ll}
\frac{1}{2} (\text{Effective Federal Funds Rate: FRED FEDFUNDS})_t & t \geq (1954:Q3) \\
\frac{1}{2} (\text{3-Month Treasury Bill: FRED TB3MS})_t & \text{else.}
\end{array} \right.
\]

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

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

\[
k_t = (1 - 0.015) k_{t-1} + \frac{\text{\(i_{t}^{eff}\)}_t}{(\text{Population above 16: FRED CNP16OV})_t}
\]

\[
i_{t}^{eff} = \omega \frac{\text{(nominal fixed investment: NIPA Table 1.1.5Q, Line 8)}_t}{(\text{Implicit price deflator fixed investment: NIPA Table 1.1.9Q, Line 8})_t} + \frac{(1 - \omega)(\text{nominal durable goods: NIPA Table 1.1.5Q, Line 4})_t}{(\text{Implicit price deflator durable goods: NIPA Table 1.1.9Q, Line 4})_t}
\]

where \(\omega\) is the average nominal share of fixed investment relative in the sum with durables.

When using an alternative definition of hours worked from Francis and Ramey (2009), I compute:

\[
n_{FR}^t = \frac{(\text{Total hours worked: Francis and Ramey (2009)})_t}{(\text{Population above 16: FRED CNP16OV})_t}
\]

Fiscal data is computed following Leeper et al. (2010), except for adding state and local governments (superscript “s&l”) to the federal government account (superscript “f”), similar to Fernandez-Villaverde et al. (2015). Since in the real world

\[
\tau^f_t = \frac{(\text{production & imports taxes: Table 3.2, Line 4})_t}{(\text{(Durables PCE)}_t + c_t) \times (\text{GDP deflator})_t - (\text{production & imports taxes})_t - (\text{Sales taxes})_t^{s&l}}
\]

\[
\tau^p_t = \frac{1}{2} (\text{(Proprietors’ income)}_t + (\text{wage income})_t + (\text{wage supplements})_t + (\text{capital income})_t)
\]

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

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

where the following NIPA sources were used:

- (Federal) production & imports taxes: Table 3.2Q, Line 4
- (State and local) sales taxes: Table 3.3Q, Line 7

A-16
• (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.

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

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

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

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 1959Q3. Since these purchases typically amount to less than 1% of federal government expenditures with a minimum of -1.1%, a maximum of 0.76%, and a median of 0.4% from 1959:Q3 to 1969:Q3, two alternative treatments of the missing data 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 was revised up. The revision covers all quarters since 2004 but not before, implying a jump in the debt time series.\(^3\) 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.

\(^3\)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).
Measured expectations and shock proxies. To control for fiscal foresight, I compile two series on the four quarter ahead federal purchases of goods and services and revenue growth from the Greenbook. To match the Greenbook data to quarters, I use the Greenbook before but closest to the middle of the second month of each quarter. This broadly matches the timing of the SPF that underlies the short-run data in Ramey (2011). It also allows me to use already digitized data on price deflators from the Real Time Data Center website at the Federal Reserve Bank of Philadelphia. Missing data is unproblematic for the defense spending forecast errors, but would be more challenging to handle in a VAR. From 1966:Q1 to 1973:Q2, some observations on three and four quarter ahead forecasts government purchases and revenue are missing. In these case, I impute them based on current and up to two quarter ahead revenue and government spending. For revenue forecasts, I additionally use Greenbook real GDP growth forecasts. I treat the imputed data as the actual data.

The above data are combined with data from Mertens and Ravn (2013) on narrative tax shock measures and new data on defense spending and monetary policy shocks constructed in the spirit of the data provided by Ramey (2011) on short-term defense spending shocks and the monetary policy shock proxy in Romer and Romer (2004).

For updating the instruments, I also use Greenbook data to update the shock series from Romer and Romer (2004). After their sample ends in 1996, I use the change in the Federal Funds Target Rate (DFEDTAR in the FRED database) to compute the desired change in the FFR rate. As in Romer and Romer (2004) I then construct the shock measure as the residual from a regression of the change in the target at an FOMC meeting on the prevailing level of the funds rate, unemployment, plus levels and changes of current and future real GDP growth and inflation. I construct inflation as the difference between the forecast for nominal and real GDP in the Greenbook. The right panels in Figure B.2 compare my updated series with the Romer and Romer (2004) series. The correlation is 0.93 over the entire sample period with observed shocks.

Ramey (2011) provides one-quarter ahead forecast errors from the Survey of Professional Forecasters (SPF) for defense spending. This series runs from 1967 to 1982. The Greenbook, in contrast, provides forecasts for defense spending on a quarterly basis since 1967. I construct the defense spending forecast error as the forecast error in the implied real defense spending growth: $E_t^GB[g_{Def,t+1}^n - π_{t+1}] - (g_{Def,t}^n - π_t)$. The left panels in Figure B.2 compare my updated series with the SPF series. The correlation is 0.84 over the entire sample period with observed shocks.
Time series comparison

Government spending proxy

Defense spending forecast error correlation = 0.84

Monetary policy proxy

Monetary shock proxy correlation = 0.93

Figure B.2: Comparing shock proxies in the literature with their updated counterparts
B.2 Approximation quality of VAR representation of DSGE model

Figure B.3: Responses of output, investment, and inflation: Quality of VAR approximation to DSGE model ($T_0^V = T_0^B \uparrow \infty$)

Note: Shown are the pointwise median and 68% and 90% posterior credible sets. Results based on lower Cholesky factorization of $S_1S'_1$. 
Figure B.4: Responses in expectations-augmented DSGE-VAR: Quality of VAR approximation
B.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 B.6 and B.7 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 (2.1) as well as of the Frobenius norm of $V$ and the log-likelihood. Figure B.5 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 B.6 and B.7). The resulting samples are also reasonably efficient: the autocorrelation of the subsamples in Figure B.5 are reasonably small, particular 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 with the informative prior and every 10th draw with the flat prior. The resulting sample is reasonably efficient also with a larger prior weight on the DSGE model.

Figure B.5: 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 \( \theta \), the plot shows 

\[
\left( \frac{1}{T/2} \sum_{s=1}^{T/2} \theta_s(i) - \frac{1}{T/2} \sum_{s=1}^{T/2} \mathbb{E}[\theta_s(i)] \right)^2 /
\left( \frac{1}{T/2} \sum_{s=1}^{T/2} \left( \theta_s(i) - \frac{1}{T/2} \sum_{u=1}^{T/2} \mathbb{E}[\theta_u(i)] \right)^2 \right)^{1/2}
\]

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 B.6: 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 $\theta$, the plot shows $t - 1 \sum_{s=1}^{\lfloor T/2 \rfloor} \theta_s(i) - 1 - 1 \sum_{s=1}^{\lfloor T/2 \rfloor} (\theta_s(i) - 1 - 1 \sum_{u=1}^{T/2} \theta_u(i))^2)^{1/2}$ 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 B.7: Brooks and Gelman (1998) type convergence diagnostic for DSGE-VAR with best-fitting model ($T_0^B = 4T$, $T_0^V = \frac{1}{5}T$)
Additional Empirical Results
– available on author’s website –

Alternative model specifications.

1. **Without cost channel.** Here I set the marginal cost equation to \( \hat{\mu}c_t = \alpha \hat{r}^d_t + (1 - \alpha) \hat{w}_t \).

2. **Monetary policy with output gap.** Here I replace the output term in (4.1) with \( \eta_{r, y} \tilde{y}_t + \eta_{r, \Delta y} \Delta \tilde{y}_t \), where \( \tilde{y}_t \) is the output gap.

3. **Monetary policy with monetary term.** Here I estimate two coefficients on government spending and debt in (4.1).

4. **Fiscal policy with monetary term.** Here I estimate a coefficient on the real interest rate in the government consumption equation (4.2a).

5. **Ad hoc investment friction.** Here I replace \( \hat{q}^b_t \) in the household FOC by \( \hat{q}^b_t + \kappa \tilde{y}_t \), where \( \kappa \) is estimated.

6. **Looser prior on noise.** In this robustness check, I increase the prior standard deviation for the relative standard deviation of measurement error from 0.5 to 1.

7. **Tighter prior on noise.** In this robustness check, I cut the prior standard deviation for the relative standard deviation of measurement error from 0.5 to 0.1.

8. **Non-diagonal G.** In this robustness check, I parametrize \( G \) as follows:

\[
G = \begin{bmatrix}
1 & 2(\kappa_{12} - 0.5) & 2(\kappa_{13} - 0.5) \\
2(\kappa_{21} - 0.5) & 1 & 2(\kappa_{23} - 0.5) \\
2(\kappa_{31} - 0.5) & 2(\kappa_{32} - 0.5) & 1
\end{bmatrix}
\times
\begin{bmatrix}
c_1 & 0 & 0 \\
0 & c_2 & 0 \\
0 & 0 & c_3
\end{bmatrix}
= \begin{bmatrix}
c_1 & 2(\kappa_{12} - 0.5)c_2 & 2(\kappa_{13} - 0.5)c_3 \\
2(\kappa_{21} - 0.5)c_1 & c_2 & 2(\kappa_{23} - 0.5)c_3 \\
2(\kappa_{31} - 0.5)c_1 & 2(\kappa_{32} - 0.5)c_2 & c_3
\end{bmatrix},
\]

where each \( \kappa \) follows a beta distribution with mean \( \frac{1}{2} \) and standard deviation \( \frac{1}{4} \). When \( \kappa = \frac{1}{2} \), the baseline case obtains. I choose this formulation to ensure that \( G \) is invertible.
(a) Variations on the structural model

(b) Variations on the measurement equation

See the previous page for details on the alternative scenarios.

Figure C.8: Robustness of marginal likelihood for varying DSGE model weights: Model specification
Figure C.9: Response of the debt-to-output ratio to the identified policy shocks

Figure C.10: Private output multipliers: Best-fitting model ($T_0^Y = \frac{1}{5} T, T_0^B = 4 \times T$).
### Weak prior

$TV_0 = T_0B = \frac{1}{2} \times T$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior mean (SD)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)-TFP</td>
<td>0.500 (0.200)</td>
<td>0.320 (0.135)</td>
</tr>
<tr>
<td>st.dev.-Transf.</td>
<td>0.500 (0.100)</td>
<td>0.332 (0.029)</td>
</tr>
<tr>
<td>st.dev.-G</td>
<td>1.000 (2.000)</td>
<td>0.417 (0.198)</td>
</tr>
<tr>
<td>AR(1)-G</td>
<td>0.500 (0.200)</td>
<td>0.577 (0.075)</td>
</tr>
<tr>
<td>st.dev.-Tax</td>
<td>0.500 (0.100)</td>
<td>0.332 (0.029)</td>
</tr>
<tr>
<td>AR(1)-Tax</td>
<td>0.500 (0.200)</td>
<td>0.320 (0.135)</td>
</tr>
<tr>
<td>st.dev.-qs</td>
<td>0.500 (0.100)</td>
<td>0.332 (0.029)</td>
</tr>
<tr>
<td>AR(1)-qs</td>
<td>0.500 (0.200)</td>
<td>0.417 (0.198)</td>
</tr>
</tbody>
</table>

### Best-fitting model

$TV_0 = \frac{1}{2} T, T_0B = 4 \times T$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)-TFP</td>
<td>0.189 (0.211)</td>
</tr>
<tr>
<td>st.dev.-Transf.</td>
<td>0.476 (0.113)</td>
</tr>
<tr>
<td>st.dev.-G</td>
<td>0.344 (0.164)</td>
</tr>
<tr>
<td>AR(1)-G</td>
<td>0.189 (0.211)</td>
</tr>
<tr>
<td>st.dev.-Tax</td>
<td>0.476 (0.113)</td>
</tr>
<tr>
<td>AR(1)-Tax</td>
<td>0.344 (0.164)</td>
</tr>
<tr>
<td>st.dev.-qs</td>
<td>0.476 (0.113)</td>
</tr>
<tr>
<td>AR(1)-qs</td>
<td>0.344 (0.164)</td>
</tr>
</tbody>
</table>

### Stronger prior

$TV_0 = T, T_0B = 4T$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)-TFP</td>
<td>0.476 (0.113)</td>
</tr>
<tr>
<td>st.dev.-Transf.</td>
<td>0.476 (0.113)</td>
</tr>
<tr>
<td>st.dev.-G</td>
<td>0.344 (0.164)</td>
</tr>
<tr>
<td>AR(1)-G</td>
<td>0.189 (0.211)</td>
</tr>
<tr>
<td>st.dev.-Tax</td>
<td>0.476 (0.113)</td>
</tr>
<tr>
<td>AR(1)-Tax</td>
<td>0.344 (0.164)</td>
</tr>
<tr>
<td>st.dev.-qs</td>
<td>0.476 (0.113)</td>
</tr>
<tr>
<td>AR(1)-qs</td>
<td>0.344 (0.164)</td>
</tr>
</tbody>
</table>

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.1: DSGE-VAR model parameter estimates with varying priors
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 in model with news shocks
Baseline model

<table>
<thead>
<tr>
<th>Flat prior</th>
<th>Best-fitting model</th>
<th>Model with news + obs. E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T_0 = 4T, T_0^V = \frac{1}{4}T)</td>
<td>(T_0 = 7.5T, T_0^V = \frac{1}{4}T)</td>
<td></td>
</tr>
</tbody>
</table>

Figure C.11: Comparison of VAR and DSGE-VAR estimated with and without expectations: Response to a government spending shock.

Table C.3: Policy rule estimates

(a) Best-fitting model: \(T_0^B = 4T, T_0^V = \frac{1}{4}T\)

<table>
<thead>
<tr>
<th>G</th>
<th>tax</th>
<th>FFR</th>
<th>Y</th>
<th>Debt</th>
<th>Inv</th>
<th>Inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>0.00 (-0.00, 0.00)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>-1.81 (-1.45, 0.27)</td>
<td>-1.41 (-4.47, 2.24)</td>
<td>0.03 (-0.30, 0.47)</td>
<td>-0.15 (-2.84, 2.44)</td>
</tr>
<tr>
<td>tax</td>
<td>0.01 (-0.24, 0.16)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>-0.30 (-1.05, 0.80)</td>
<td>-0.33 (-0.94, 0.54)</td>
<td>0.00 (-0.15, 0.16)</td>
<td>-0.48 (-1.50, 0.63)</td>
</tr>
<tr>
<td>FFR</td>
<td>-0.02 (-0.05, 0.00)</td>
<td>0.04 (-0.03, 0.12)</td>
<td>0.05 (-0.07, 0.19)</td>
<td>0.03 (-0.08, 0.14)</td>
<td>-0.01 (-0.03, 0.01)</td>
<td>0.07 (-0.08, 0.22)</td>
</tr>
</tbody>
</table>

(b) Stronger prior: \(T_0^B = 4T, T_0^V = T\)

<table>
<thead>
<tr>
<th>G</th>
<th>tax</th>
<th>FFR</th>
<th>Y</th>
<th>Debt</th>
<th>Inv</th>
<th>Inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>-0.00 (-0.00, 0.00)</td>
<td>-0.00 (-0.00, 0.00)</td>
<td>-1.17 (-1.04, 0.02)</td>
<td>-0.90 (-2.08, 0.01)</td>
<td>0.03 (-0.16, 0.27)</td>
<td>0.04 (-0.25, 1.28)</td>
</tr>
<tr>
<td>tax</td>
<td>0.01 (-0.11, 0.11)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>-0.08 (-0.52, 0.44)</td>
<td>-0.10 (-0.44, 0.28)</td>
<td>0.00 (-0.08, 0.09)</td>
<td>-0.11 (-0.63, 0.38)</td>
</tr>
<tr>
<td>FFR</td>
<td>-0.02 (-0.04, 0.01)</td>
<td>0.04 (-0.01, 0.08)</td>
<td>0.07 (-0.02, 0.16)</td>
<td>0.04 (-0.03, 0.11)</td>
<td>-0.01 (-0.02, 0.01)</td>
<td>0.13 (0.03, 0.23)</td>
</tr>
</tbody>
</table>

(c) Dogmatic prior: \(T_0^B = T_0^V = \infty\)

<table>
<thead>
<tr>
<th>G</th>
<th>tax</th>
<th>FFR</th>
<th>Y</th>
<th>Debt</th>
<th>Inv</th>
<th>Inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>-0.00 (-0.00, 0.00)</td>
<td>-0.00 (-0.00, 0.00)</td>
<td>-0.00 (-0.00, -0.00)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>-0.00 (-0.00, 0.00)</td>
<td>0.23 (0.20, 0.27)</td>
</tr>
<tr>
<td>tax</td>
<td>-0.00 (-0.00, -0.00)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>0.01 (0.01, 0.02)</td>
<td>-0.00 (-0.00, -0.00)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>-0.00 (-0.00, 0.00)</td>
</tr>
<tr>
<td>FFR</td>
<td>-0.00 (-0.00, -0.00)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>0.02 (0.01, 0.03)</td>
<td>-0.00 (-0.00, -0.00)</td>
<td>0.00 (-0.00, 0.00)</td>
<td>-0.00 (-0.00, 0.00)</td>
</tr>
</tbody>
</table>

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 dispersed and are bounded away from zero only with a very strong prior.
Figure C.12: Comparison of VAR and DSGE-VAR estimated with and without expectations: Response to a monetary policy shock.