Supplement to "A narrative approach to a fiscal DSGE model"

(*Quantitative Economics*, Vol. 11, No. 2, May 2020, 801–837)

Thorsten Drautzburg
Federal Reserve Bank of Philadelphia

This appendix provides proofs and additional details on the model in part A. Part B describes the data and additional empirical results.

Appendix 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

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

so that $\mathbf{A}_{21} = \boldsymbol{\kappa}\mathbf{A}_{11}$. Then

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

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

$$\mathbf{u}_t = \mathbf{A}\boldsymbol{\varepsilon}_t = \begin{bmatrix} \mathbf{A}^{[1]} & \mathbf{A}^{[2]} \end{bmatrix} \boldsymbol{\varepsilon}_t = \mathbf{A}^{[1]}\boldsymbol{\varepsilon}_t^{[1]} + \mathbf{A}^{[2]}\boldsymbol{\varepsilon}_t^{[2]} = \begin{bmatrix} \mathbf{I}_{m_z} \\ \boldsymbol{\kappa} \end{bmatrix} \mathbf{A}_{11}\boldsymbol{\varepsilon}_t^{[1]} + \begin{bmatrix} \mathbf{A}_{12} \\ \mathbf{A}_{22} \end{bmatrix} \boldsymbol{\varepsilon}_t^{[2]}.$$

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

$$\mathrm{Var}\left[\mathbf{u}_t | \boldsymbol{\varepsilon}_t^{[1]}\right] = \mathbf{A}^{[2]}\left(\mathbf{A}^{[2]}\right)' = \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{12}' & \mathbf{A}_{12}\mathbf{A}_{22}' \\ \mathbf{A}_{22}\mathbf{A}_{12}' & \mathbf{A}_{22}\mathbf{A}_{22}' \end{bmatrix},$$

$$\mathrm{Var}\left[\mathbf{u}_t | \boldsymbol{\varepsilon}_t^{[2]}\right] = \mathbf{A}^{[1]}\left(\mathbf{A}^{[1]}\right)' = \begin{bmatrix} \mathbf{A}_{11}\mathbf{A}_{11}' & \mathbf{A}_{11}\mathbf{A}_{11}\boldsymbol{\kappa} \\ \boldsymbol{\kappa}\mathbf{A}_{11}\mathbf{A}_{11}' & \boldsymbol{\kappa}\mathbf{A}_{11}\mathbf{A}_{11}'\boldsymbol{\kappa} \end{bmatrix},$$

$$\Sigma = \mathrm{Var}[\mathbf{u}_t] = \mathrm{Var}\left[\mathbf{u}_t | \boldsymbol{\varepsilon}_t^{[2]}\right] + \mathrm{Var}\left[\mathbf{u}_t | \boldsymbol{\varepsilon}_t^{[1]}\right] = \begin{bmatrix} \Sigma_{12}\Sigma_{12}' & \Sigma_{12}\Sigma_{22}' \\ \Sigma_{22}\Sigma_{12}' & \Sigma_{22}\Sigma_{22}' \end{bmatrix}.$$

Thorsten Drautzburg: tdrautzburg@gmail.com

 https://doi.org/10.3982/QE1083

Note that

$$\mathbf{u}_t^{res} \equiv \mathbf{u}_t - \mathbb{E}\big[\mathbf{u}_t|\boldsymbol{\varepsilon}_t^{[1]}\big] \perp \mathbb{E}\big[\mathbf{u}_t|\boldsymbol{\varepsilon}_t^{[1]}\big] = \begin{bmatrix} \mathbf{I}_{m_z} \\ \boldsymbol{\kappa} \end{bmatrix} \mathbf{A}_{11}\boldsymbol{\varepsilon}_t^{[1]}.$$

Any vector in the nullspace of $[\mathbf{I}_{m_z}\ \boldsymbol{\kappa}']$ satisfies the orthogonality condition.

Note that $\{[\begin{smallmatrix}\mathbf{I}_{m_z}\\ \boldsymbol{\kappa}\end{smallmatrix}], [\begin{smallmatrix}\boldsymbol{\kappa}'\\ -\mathbf{I}_{m-m_z}\end{smallmatrix}]\}$ is an orthogonal basis for $\mathbb{R}^m$.

Define

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

Note that $\mathbf{Z}^{[2]}$ spans the nullspace of $\mathbf{A}^{[1]'}$. Hence, $(\mathbf{Z}^{[2]})'\mathbf{v}_t$ projects $\mathbf{v}_t$ onto the nullspace of the instrument-identified shocks $\boldsymbol{\varepsilon}_t^{[1]}$.

$$\begin{aligned}(\mathbf{Z}^{[2]})'\mathbf{v}_t &= (\mathbf{Z}^{[2]})'\mathbf{A}\boldsymbol{\varepsilon}_t = (\mathbf{Z}^{[2]})'\begin{bmatrix}\mathbf{A}^{[1]} & \mathbf{A}^{[2]}\end{bmatrix}\epsilon_t = (\mathbf{Z}^{[2]})'\begin{bmatrix}\mathbf{Z}^{[1]}\|\tilde{\mathbf{Z}}\|\mathbf{A}_{11} & \mathbf{A}^{[2]}\end{bmatrix}\boldsymbol{\varepsilon}_t \\ &= \begin{bmatrix}\mathbf{0} & (\mathbf{Z}^{[2]})'\mathbf{A}^{[2]}\end{bmatrix}\boldsymbol{\varepsilon}_t = \mathbf{0}\times\boldsymbol{\varepsilon}_t^{[1]} + (\mathbf{Z}^{[2]})'\mathbf{A}^{[2]}\boldsymbol{\varepsilon}_t^{[2]} \perp\!\!\!\perp \boldsymbol{\varepsilon}^{[1]}.\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\big[\mathbf{v}_t|\boldsymbol{\varepsilon}_t^{[2]}\big] &= \mathrm{Cov}\big[\mathbf{v}_t, (\mathbf{Z}^{[2]})'\mathbf{v}_t\big]\,\mathrm{Var}\big[(\mathbf{Z}^{[2]})'\mathbf{v}_t\big]^{-1}(\mathbf{Z}^{[2]})'\mathbf{v}_t, \\
\mathbf{v}_t - \mathbb{E}\big[\mathbf{v}_t|\boldsymbol{\varepsilon}_t^{[2]}\big] &= \big(\mathbf{I} - \mathrm{Cov}\big[\mathbf{v}_t, (\mathbf{Z}^{[2]})'\mathbf{v}_t\big]\,\mathrm{Var}\big[(\mathbf{Z}^{[2]})'\mathbf{v}_t\big]^{-1}(\mathbf{Z}^{[2]})'\big)\mathbf{v}_t, \\
\mathrm{Cov}\big[\mathbf{v}_t, (\mathbf{Z}^{[2]})'\mathbf{v}_t\big] &= \boldsymbol{\Sigma}\mathbf{Z}^{[2]} = \boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z}\end{bmatrix}, \\
\mathrm{Var}\big[\mathbf{v}_t|\boldsymbol{\varepsilon}_t^{[2]}\big] &= \mathbb{E}\big[\big(\mathbf{I} - \mathrm{Cov}\big[\mathbf{v}_t, (\mathbf{Z}^{[2]})'\mathbf{v}_t\big]\,\mathrm{Var}\big[(\mathbf{Z}^{[2]})'\mathbf{v}_t\big]^{-1}(\mathbf{Z}^{[2]})'\big)\mathbf{v}_t\mathbf{v}_t'\big] \\
&= \mathbb{E}\big[\mathbf{v}_t\mathbf{v}_t'\big] - \mathrm{Cov}\big[\mathbf{v}_t, (\mathbf{Z}^{[2]})'\mathbf{v}_t\big]\,\mathrm{Var}\big[(\mathbf{Z}^{[2]})'\mathbf{v}_t\big]^{-1}\mathbb{E}\big[(\mathbf{Z}^{[2]})'\big)\mathbf{v}_t\mathbf{v}_t'\big] \\
&= \boldsymbol{\Sigma} - \mathrm{Cov}\big[\mathbf{v}_t, (\mathbf{Z}^{[2]})'\mathbf{v}_t\big]\,\mathrm{Var}\big[(\mathbf{Z}^{[2]})'\mathbf{v}_t\big]^{-1}\,\mathrm{Cov}\big[\mathbf{v}_t, (\mathbf{Z}^{[2]})'\mathbf{v}_t\big] \\
&= \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z}\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{\kappa} & -\mathbf{I}_{m-m_z}\end{bmatrix}\boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{\kappa} & -\mathbf{I}_{m-m_z}\end{bmatrix}\Sigma \\
&= \begin{bmatrix}\mathbf{A}_{11}\mathbf{A}_{11}' & \mathbf{A}_{11}\mathbf{A}_{11}\boldsymbol{\kappa} \\ \boldsymbol{\kappa}\mathbf{A}_{11}\mathbf{A}_{11}' & \boldsymbol{\kappa}\mathbf{A}_{11}\mathbf{A}_{11}'\boldsymbol{\kappa}\end{bmatrix}.
\end{aligned} \tag{A.4}$$

This gives a solution for $\mathbf{A}_{11}\mathbf{A}_{11}'$ in terms of observables: $\boldsymbol{\Sigma}$ and $\boldsymbol{\kappa} = \boldsymbol{\Gamma}_1^{-1}\times\boldsymbol{\Gamma}_2$. For future reference, note that this also implies that

$$\begin{aligned}
\mathrm{Var}\big[\mathbf{v}_t|\boldsymbol{\varepsilon}_t^{[1]}\big] &= \boldsymbol{\Sigma} - \mathrm{Var}\big[\mathbf{v}_t|\boldsymbol{\varepsilon}_t^{[2]}\big] \\
&= \boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z}\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{\kappa} & -\mathbf{I}_{m-m_z}\end{bmatrix}\boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z}\end{bmatrix}'\boldsymbol{\Sigma}.
\end{aligned} \tag{A.5}$$

In general, $\mathbf{A}_{11}$ 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 $\mathbf{A}_{11}\mathbf{A}_{11}'$. Given $\mathbf{A}_{11}$, the impact response to a unit shock is given by

$$\begin{bmatrix} \mathbf{I}_{m_z} \\ \boldsymbol{\kappa} \end{bmatrix} \mathbf{A}_{11}.$$

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 $\mathbf{v}_t$ in terms of structural shocks $\boldsymbol{\varepsilon}_t$:

$$\mathbf{v}_t = \mathbf{A}\boldsymbol{\varepsilon}_t \equiv \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1}, \qquad \mathbf{v}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \end{bmatrix}. \tag{A.6}$$

Note that

$$\begin{aligned} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} &= \begin{bmatrix} \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & -\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \end{bmatrix}. \end{aligned}$$

Note that

$$\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} = \mathbf{A}_{11}^{-1}\left(\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)\mathbf{A}_{11}^{-1}\right)^{-1} = \mathbf{A}_{11}^{-1}\left(I - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\right)^{-1}$$

and define

$$\mathbf{S}_1 \equiv \left(\mathbf{I} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\right)\mathbf{A}_{11}, \qquad \mathbf{S}_2 \equiv \left(\mathbf{I} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\right)\mathbf{A}_{22} \tag{A.7}$$

so that

$$\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} = \mathbf{S}_1^{-1}, \qquad \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} = \mathbf{S}_2^{-1}.$$

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

$$\begin{aligned} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} \mathbf{v}_t &= \begin{bmatrix} \mathbf{S}_1^{-1} & -\mathbf{S}_1^{-1}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ -\mathbf{S}_2^{-1}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{S}_2^{-1} \end{bmatrix} \mathbf{v}_t \\ &= \begin{bmatrix} \mathbf{S}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \mathbf{v}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \end{bmatrix} \end{aligned}$$

and equivalently

$$\begin{bmatrix} \mathbf{I} & -\boldsymbol{\eta} \\ -\boldsymbol{\kappa} & \mathbf{I} \end{bmatrix} \mathbf{v}_t = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \end{bmatrix} \tag{A.8}$$

defining $\boldsymbol{\eta} \equiv \mathbf{A}_{12}\mathbf{A}_{22}^{-1}$ and $\boldsymbol{\kappa} \equiv \mathbf{A}_{21}\mathbf{A}_{11}^{-1}$. Equation (3.4) follows immediately.

LEMMA 1 (Mertens and Ravn (2013)). *Let* $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$ *and* $\boldsymbol{\Gamma} = [\mathbf{G}\ \mathbf{0}]\mathbf{A}'$, *where* $\mathbf{G}$ *is an* $m_z \times m_z$ *invertible matrix and* $\mathbf{A}$ *is of full rank. Then* $\mathbf{A}^{[1]}$ *is identified up to a factorization of* $\mathbf{S}_1\mathbf{S}_1'$ *with* $\mathbf{S}_1$ *defined in* (A.7).

PROOF. Since $\mathbf{A}$ is of full rank, it is invertible and (A.8) holds for any such $\mathbf{A}$. Given $\boldsymbol{\eta}$, $\boldsymbol{\kappa}$, (A.8) implies (3.5), which I reproduce here for convenience:

$$\mathbf{A}^{[1]} = \begin{bmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{21} \end{bmatrix} = \begin{bmatrix} (\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1} \\ (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta})^{-1}\boldsymbol{\kappa} \end{bmatrix} \operatorname{chol}\big(\mathbf{S}_1\mathbf{S}_1'\big). \tag{3.5}$$

If $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ pin down $\boldsymbol{\eta}$, $\boldsymbol{\kappa}$ uniquely, $\mathbf{A}^{[1]}$ is uniquely identified except for a factorization of $\mathbf{S}_1\mathbf{S}_1'$.

To show that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ pin down $\boldsymbol{\eta}$, $\boldsymbol{\kappa}$ uniquely, consider $\boldsymbol{\kappa}$ first. Since $\boldsymbol{\Gamma} = [\mathbf{G}\ \mathbf{0}]\mathbf{A}$ and $\mathbf{G}$ is an $m_z \times m_z$ invertible matrix, it follows that Assumption 1 holds. It then follows from (3.2) that $\boldsymbol{\kappa} = \mathbf{A}_{21}\mathbf{A}_{11}^{-1} = \boldsymbol{\Gamma}^2\boldsymbol{\Gamma}_1^{-1}$.

To compute $\boldsymbol{\eta}$, more algebra is needed. Partition $\boldsymbol{\Sigma} = \left[\begin{smallmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{smallmatrix}\right]$, where $\boldsymbol{\Sigma}_{11}$ is $m_z \times m_z$, $\boldsymbol{\Sigma}_{12}$ is $m_z \times (m - m_z)$ and $\boldsymbol{\Sigma}_{22}$ is $(m - m_z) \times (m - m_z)$. Define

$$\mathbf{A}_{22}\mathbf{A}_{22}' = \boldsymbol{\Sigma}_{22} - \boldsymbol{\kappa}\mathbf{A}_{11}\mathbf{A}_{11}'\boldsymbol{\kappa}' = \boldsymbol{\Sigma}_{22} - \boldsymbol{\kappa}\big(\boldsymbol{\Sigma}_{11} - \mathbf{A}_{12}\mathbf{A}_{12}'\big)\boldsymbol{\kappa}',$$

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

$$\mathbf{A}_{12}\mathbf{A}_{12}' = \big(\boldsymbol{\Sigma}_{12}' - \boldsymbol{\kappa}\boldsymbol{\Sigma}_{11}\big)'\big(\mathbf{Z}\mathbf{Z}'\big)^{-1}\big(\boldsymbol{\Sigma}_{12}' - \boldsymbol{\kappa}\boldsymbol{\Sigma}_{11}\big)$$

with

$$\mathbf{Z}\mathbf{Z}' = \boldsymbol{\kappa}\boldsymbol{\Sigma}_{11}\boldsymbol{\kappa}' - \big(\boldsymbol{\Sigma}_{12}'\boldsymbol{\kappa}' + \boldsymbol{\kappa}\boldsymbol{\Sigma}_{12}\big) + \boldsymbol{\Sigma}_{22} = \begin{bmatrix} \boldsymbol{\kappa} & -\mathbf{I}_{m-m_z} \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z} \end{bmatrix}.$$

The coefficient matrix of interest, $\boldsymbol{\eta}$, is then defined as

$$\begin{aligned} \boldsymbol{\eta} \equiv \mathbf{A}_{12}\mathbf{A}_{22}^{-1} = \mathbf{A}_{12}\mathbf{A}_{22}'\big(\mathbf{A}_{22}\mathbf{A}_{22}'\big)^{-1} &= \big(\boldsymbol{\Sigma}_{12} - \boldsymbol{\kappa}\mathbf{A}_{11}\mathbf{A}_{11}'\big)'\big(\mathbf{A}_{22}\mathbf{A}_{22}'\big)^{-1} \\ &= \big(\boldsymbol{\Sigma}_{12} - \boldsymbol{\kappa}\boldsymbol{\Sigma}_{11}' + \boldsymbol{\kappa}\mathbf{A}_{12}\mathbf{A}_{12}'\big)'\big(\mathbf{A}_{22}\mathbf{A}_{22}'\big)^{-1}. \end{aligned}$$

Thus, $\boldsymbol{\eta}$ and $\boldsymbol{\kappa}$ are uniquely identified given $\boldsymbol{\Sigma}$, $\boldsymbol{\Gamma}$. □

The above derivations link $\mathbf{S}_1$ to $\mathbf{A}^{-1}$. I now compute $\mathbf{S}_1$ for a class of models.

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

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

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

Proof. Given Lemma 1, $\mathbf{A}^{[1]}$ is identified uniquely if $\mathbf{S}_1$ is identified uniquely. In what follows, I establish that under the ordering in the proposition, $\mathbf{S}_1$, as defined in (A.7) for arbitrary full rank $\mathbf{A}$, is unique up to a normalization. It then follows that $\mathbf{A}^{[1]}$ is identified uniquely, and hence, equal to $\mathbf{A}^*[\mathbf{I}_{m_z}, \mathbf{0}_{(m-m_z)\times(m-m_z)}]'$.

To proceed, stack the m_p policy rules:

$$
\begin{aligned}
\mathbf{y}_t^p &= \sum_{i=m_p+1}^{m} \begin{bmatrix} \eta_{1,i} & 0 & \dots & 0 \\ 0 & \eta_{2,i} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \eta_{m_p,i} \end{bmatrix} y_{i,t} + \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \vdots \\ \boldsymbol{\lambda}_{n_p} \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} \boldsymbol{\Sigma}_{11} & 0 & \dots & 0 \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ {}_{n_p,1} & \sigma_{n_p,2} & \dots & \sigma_{n_p,n_p} \end{bmatrix} \boldsymbol{\epsilon}_t^p \\
&\equiv \sum_{i=m_p+1}^{m} \mathbf{D}_i y_{i,t} + \boldsymbol{\Lambda}\mathbf{x}_{t-1} + \begin{bmatrix} \mathbf{D}_0 & \mathbf{0} \end{bmatrix} \boldsymbol{\varepsilon}_t, \\
&= \left(\begin{bmatrix} \mathbf{D}_0 & \mathbf{0} \end{bmatrix} \boldsymbol{\varepsilon}_t + \sum_{i=m_p+1}^{m} \mathbf{D}_i \mathbf{1}\mathbf{A}_i' \right) \boldsymbol{\varepsilon}_t + \left(\sum_{i=m_p+1}^{m} \mathbf{D}_i \mathbf{1}\mathbf{B}_i' \mathbf{x}_{t-1} + \boldsymbol{\Lambda} \right) \mathbf{x}_{t-1},
\end{aligned}
$$

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

Without loss of generality, order the policy instruments first, before the $m - m_p = \bar{n}$ nonpolicy variables. Then $\mathbf{A}^*$ in the observation equation (2.2a) can be written as

$$
\begin{bmatrix} [\mathbf{D}_0, \mathbf{0}] + \sum_{i=m_p}^{m} \mathbf{D}_i \mathbf{1} (\mathbf{A}_i^*)' \\ (\mathbf{A}_{m_p+1}^*)' \\ \vdots \\ (\mathbf{A}_m^*)' \end{bmatrix},
$$

where $\underline{0}$ is a full-rank lower diagonal matrix and the $\mathbf{D}_j$ matrices are $m_p \times m_p$ matrices.

To find $(\mathbf{A}^*)^{-1}$, proceed by Gauss–Jordan elimination to rewrite the system $\mathbf{A}^*\mathbf{X} = \mathbf{I}_m$, with solution $\mathbf{X} = (\mathbf{A}^*)^{-1}$, as $[\mathbf{A}^*|\mathbf{I}_m]$. Define $\mathbf{E}$ as a conformable matrix such that

$[\mathbf{A}^*|\mathbf{I}_m] \overset{\mathbf{E}}{\to} [\mathbf{B}|\mathbf{C}] = [\mathbf{E}\mathbf{A}^*|\mathbf{E}\mathbf{I}_m]$. Then

$$\left[(\mathbf{A}^*)|\mathbf{I}_m\right] = \left[\begin{array}{c|ccccc} \begin{bmatrix}\mathbf{D}_0 & \mathbf{0}\end{bmatrix} + \sum_{i=m_p+1}^{m} \mathbf{D}_i \mathbf{1}(\mathbf{A}_i^*)' & \mathbf{I}_{m_p} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ (\mathbf{A}_{m_p+1}^*)' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (\mathbf{A}_m^*)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right]$$

$$\overset{\mathbf{E}_1}{\to} \left[\begin{array}{c|ccccc} \begin{bmatrix}\mathbf{D}_0 & \mathbf{0}\end{bmatrix} + \sum_{i=m_p+2}^{m} \mathbf{D}_i \mathbf{1}(A_i^*)' & \mathbf{I}_{m_p} & -\mathbf{D}_{m_p+1}\mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ (\mathbf{A}_{m_p+1}^*)' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (\mathbf{A}_{m_p+2}^*)' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (\mathbf{A}_m^*)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right]$$

$$\overset{\mathbf{E}_2}{\to} \left[\begin{array}{c|ccccc} \begin{bmatrix}\mathbf{D}_0 & \mathbf{0}\end{bmatrix} + \sum_{i=m_p+3}^{m} \mathbf{D}_i \mathbf{1}(A_i^*)' & \mathbf{I}_{m_p} & -\mathbf{D}_{m_p+1}\mathbf{1} & -\mathbf{D}_{m_p+2}\mathbf{1} & \dots & \mathbf{0} \\ (\mathbf{A}_{m_p+1}^*)' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (\mathbf{A}_{m_p+2}^*)' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (\mathbf{A}_m*)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right] \overset{E_3}{\to} \dots$$

$$\overset{\mathbf{E}_{m-m_p}}{\to} \left[\begin{array}{c|ccccc} \begin{bmatrix}\mathbf{D}_0 & \mathbf{0}\end{bmatrix} & \mathbf{I}_{m_p} & -\mathbf{D}_{m_p+1}\mathbf{1} & -\mathbf{D}_{m_p+2}\mathbf{1} & \dots & -\mathbf{D}_m\mathbf{1} \\ (\mathbf{A}_{m_p+1}^*)' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (\mathbf{A}_{m_p+2}^*)' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (\mathbf{A}_m*)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right]$$

$$\overset{\mathbf{E}_D}{\to} \left[\begin{array}{c|ccccc} \begin{bmatrix}\mathbf{I}_{m_p} & \mathbf{0}\end{bmatrix} & \mathbf{D}_0^{-1} & -\mathbf{D}_0^{-1}\mathbf{D}_{m_p+1}\mathbf{1} & -\mathbf{D}_0^{-1}\mathbf{D}_{m_p+2}\mathbf{1} & \dots & -\mathbf{D}_0^{-1}\mathbf{D}_m\mathbf{1} \\ (\mathbf{A}_{m_p+1}^*)' & \mathbf{0}' & 1 & 0 & \dots & 0 \\ (\mathbf{A}_{m_p+2}^*)' & \mathbf{0}' & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (\mathbf{A}_m*)' & \mathbf{0}' & 0 & 0 & \dots & 1 \end{array}\right].$$

Thus, $((\mathbf{A}^*)^{-1})_{1:m_p,1:m_p} = (\mathbf{E}_D\, \mathbf{E}_{m-m_p}\, \dots\, \mathbf{E}_1\, I_m)_{1:m_p,1:m_p}$.

Now consider cases (a) and (b):

(a) $m_z = m_p$. From (A.7), $\mathbf{S}_1$ is the upper left corner of $(\mathbf{A}^*)^{-1}$:

$$\mathbf{S}_1 \equiv \left((\mathbf{A}^*)^{-1}\right)_{1:m_p,1:m_p} = \mathbf{D}_0^{-1}$$

and $\mathbf{S}_1$ is a (lower) diagonal matrix because $\mathbf{D}_0$ is (lower) diagonal.

(b) $m_z = m_p + 1$, $\eta_{p,m_p+1} = 0$, $p = 1, \ldots, m_p$. The second condition implies that $\mathbf{D}_{m_p+1} = \mathbf{0}_{m_p \times m_p}$. It follows that $\mathbf{S}_1$ defined in (A.7) is given by

$$\mathbf{S}_1 \equiv \left((\mathbf{A}^*)^{-1}\right)_{1:m_p+1,1:m_p+1} = \begin{bmatrix} \mathbf{D}_0^{-1} & \mathbf{D}_{m_p+1}\mathbf{1} \\ s_{m_p+1,1:m_p} & s_{m_p+1,m_p+1} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_0^{-1} & \mathbf{0} \\ s_{m_p+1,1:m_p} & s_{m_p+1,m_p+1} \end{bmatrix}.$$

Thus, $\mathbf{S}_1$ is lower triangular.

In both cases, $\mathbf{S}_1$ is lower triangular. Since the lower Cholesky decomposition is unique, a Cholesky decomposition of $\mathbf{S}_1\mathbf{S}_1'$ recovers $\mathbf{S}_1$ if we normalize signs of the diagonal of $\mathbf{S}_1$ to be positive. Given identification of $\mathbf{S}_1$, the identification of $\mathbf{A}^{[1]}$ follows from Lemma 1. □

A.3 *VAR priors and posteriors*

Let $\mathbf{u}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{V})$ and let $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_T]'$, where $\mathbf{u}_t$ is $m_a \times 1$ and $\mathbf{U}$ is $T \times m_a$. Then the likelihood can be written as

$$\begin{aligned} L &= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \mathbf{u}_t'\mathbf{V}^{-1}\mathbf{u}_t\right) \\ &= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \operatorname{tr}\left(\mathbf{u}_t'\mathbf{V}^{-1}\mathbf{u}_t\right)\right) \\ &= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\mathbf{V}^{-1}\sum_{t=1}^{T} \mathbf{u}_t\mathbf{u}_t'\right)\right) \\ &= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\mathbf{V}^{-1}\mathbf{U}'\mathbf{U}\right)\right) \qquad \text{(A.9)} \\ &= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{vec}(\mathbf{U})'\left(\mathbf{V}^{-1} \otimes \mathbf{I}_T\right)\operatorname{vec}(\mathbf{U})\right), \end{aligned}$$

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

For the SUR model, $[\mathbf{Y}, \mathbf{Z}] = [\mathbf{X}_y, \mathbf{X}_z]\left[\begin{smallmatrix}\mathbf{B}_y \\ \mathbf{B}_z\end{smallmatrix}\right] + \mathbf{U}$. Consequently, $\mathbf{Y}_{\text{SUR}} \equiv \operatorname{vec}([\mathbf{Y}, \mathbf{Z}]) = \mathbf{X}_{\text{SUR}} \operatorname{vec}\left(\left[\begin{smallmatrix}\mathbf{B}_y \\ \mathbf{B}_z\end{smallmatrix}\right]\right) + \operatorname{vec}(\mathbf{U})$, where

$$\mathbf{X}_{\text{SUR}} \equiv \begin{bmatrix} \mathbf{I}_m \otimes \mathbf{X}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_z} \otimes \mathbf{X}_z \end{bmatrix}.$$

The likelihood can then also be written as

$$\begin{aligned} L &= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2} \exp\left(-\frac{1}{2}(\mathbf{Y}_{\text{SUR}} - \mathbf{X}_{\text{SUR}}\beta)'\left(\mathbf{V}^{-1} \otimes \mathbf{I}_T\right)(\mathbf{Y}_{\text{SUR}} - \mathbf{X}_{\text{SUR}}\boldsymbol{\beta})\right) \\ &= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2} \exp\left(-\frac{1}{2}(\tilde{\mathbf{Y}}_{\text{SUR}} - \tilde{\mathbf{X}}_{\text{SUR}}\boldsymbol{\beta})'(\tilde{\mathbf{Y}}_{\text{SUR}} - \tilde{\mathbf{X}}_{\text{SUR}}\boldsymbol{\beta})\right) \end{aligned}$$

$$= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2}\exp\Big(-\frac{1}{2}(\tilde{\mathbf{Y}}_{\text{SUR}} - \tilde{\mathbf{X}}_{\text{SUR}}\boldsymbol{\beta})'(\tilde{\mathbf{Y}}_{\text{SUR}} - \tilde{\mathbf{X}}_{\text{SUR}}\boldsymbol{\beta})\Big)$$

$$= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2}\exp\Big(-\frac{1}{2}\big(\tilde{\mathbf{X}}_{\text{SUR}}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_{\text{SUR}})\big)'\big(\tilde{\mathbf{X}}_{\text{SUR}}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_{\text{SUR}})\big)\Big)$$

$$= (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2}\exp\Big(-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_{\text{SUR}})'(\tilde{\mathbf{X}}'_{\text{SUR}}\tilde{\mathbf{X}}_{\text{SUR}})(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_{\text{SUR}})\Big), \tag{A.10}$$

where $\tilde{\boldsymbol{\beta}}_{\text{SUR}} \equiv (\tilde{\mathbf{X}}'_{\text{SUR}}\tilde{\mathbf{X}}_{\text{SUR}})^{-1}\tilde{\mathbf{X}}'_{\text{SUR}}\tilde{\mathbf{Y}}_{\text{SUR}}$ 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 $\boldsymbol{\beta}$: $\boldsymbol{\beta}|\mathbf{V}^{-1} \sim \mathcal{N}(\tilde{\boldsymbol{\beta}}_{\text{SUR}}, (\tilde{\mathbf{X}}'_{\text{SUR}}\tilde{\mathbf{X}}_{\text{SUR}})^{-1}) \equiv \mathcal{N}(\tilde{\boldsymbol{\beta}}_{\text{SUR}}, (\mathbf{X}'_{\text{SUR}}(\mathbf{V}^{-1} \otimes \mathbf{I})\mathbf{X}_{\text{SUR}})^{-1})$. Alternatively, expression (A.9) for the likelihood is proportional to a conditional Wishart distribution for $\mathbf{V}^{-1}$: $\mathbf{V}^{-1}|\boldsymbol{\beta} \sim \mathcal{W}_{m_a}((\mathbf{U}(\boldsymbol{\beta})'\mathbf{U}(\boldsymbol{\beta}))^{-1}, T + m_a + 1)$. Premultiplying with a Jeffrey's prior over $\mathbf{V}$, transformed to $\mathbf{V}^{-1}$, is equivalent to premultiplying by $\pi(\mathbf{V}^{-1}) \equiv |\mathbf{V}^{-1}|^{-\frac{m_a+1}{2}}$ and yields

$$\pi\big(\mathbf{V}^{-1}\big)\times = \big|\mathbf{V}^{-1}\big|^{-\frac{m_a+1}{2}} \times (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2}\exp\Big(-\frac{1}{2}\operatorname{tr}\big(\mathbf{V}^{-1}\mathbf{U}'\mathbf{U}\big)\Big)$$

$$= (2\pi)^{-mT/2}\big|\mathbf{V}^{-1}\big|^{(T-m_a-1)/2}\exp\Big(-\frac{1}{2}\operatorname{tr}\big(\mathbf{V}^{-1}\mathbf{U}'\mathbf{U}\big)\Big), \tag{A.11}$$

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

$$\text{SSR}(\beta) \equiv \mathbf{U}(\boldsymbol{\beta})'\mathbf{U}(\boldsymbol{\beta}) = \big[\mathbf{Y} - \mathbf{X}_y\mathbf{B}_y(\boldsymbol{\beta}), \mathbf{Z} - \mathbf{X}_z\mathbf{B}_z(\boldsymbol{\beta})\big]'\big[\mathbf{Y} - \mathbf{X}_y\mathbf{B}_y(\boldsymbol{\beta}), \mathbf{Z} - \mathbf{X}_z\mathbf{B}_z(\boldsymbol{\beta})\big]$$

$$= \sum_{t=1}^{T}\big[\mathbf{y}_t - \mathbf{x}_{y,t}\mathbf{B}_y(\boldsymbol{\beta}), \mathbf{z}_t - \mathbf{x}_{z,t}\mathbf{B}_z(\boldsymbol{\beta})\big]'\big[\mathbf{y}_t - \mathbf{x}_{y,t}\mathbf{B}_y(\beta), \mathbf{z}_t - \mathbf{x}_{z,t}\mathbf{B}_z(\boldsymbol{\beta})\big].$$

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, Rubio-Ramirez, Sargent, and Watson (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. $\mathbf{A}^*$ *is nonsingular, and the matrix* $\mathbf{C}^* - \mathbf{D}^*(\mathbf{A}^*)^{-1}\mathbf{B}^*$ *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.[1]

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

LEMMA 2 (Fernandez-Villaverde et al. (2007, p. 1022)). *Let* $\mathbf{y}_t$ *be generated by the DSGE economy* (2.2). *Under Assumption* 1, *the variance-covariance matrix of the one-step-ahead prediction error in the Wold representation of* $\mathbf{y}_t$ *is given by* $\boldsymbol{\Sigma}^* = (\mathbf{A}^*)(\mathbf{A}^*)'$.

In my application, I assume throughout that Assumption 1 holds so that a VAR(p) can approximate the DSGE model dynamics arbitrarily well. Thus, $\mathbf{A}\mathbf{A}' \approx \mathbf{A}^*(\mathbf{A}^*)'$. This assumption is not necessarily satisfied and, in general, depends on the observables $\mathbf{y}_t$.[2] With an equal number of AR(1) shock processes as observables, I found two intuitive cases in my exploratory analysis that violate Assumption 1 for most of the parameter space: First, a model with capital that exclude investment and capital as observables. This is similar to Chari, Kehoe, and McGrattan (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 $\mathcal{N}(\boldsymbol{\beta}_0, \bar{\mathbf{V}}_0^{-1})$ distribution, whereas the observations that generate the prior for the covariance matrix are generated from a $\mathcal{N}(\mathbf{0}, \mathbf{V}^{-1})$ distribution.

Specify

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

Note that this is not equal to

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

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

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

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

Note that because the prior for $\boldsymbol{\beta}$ 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

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

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

The joint density is given by

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

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

$$p\big(\mathbf{Y},\mathbf{Z}|\boldsymbol{\beta},\mathbf{V}^{-1}\big) = (2\pi)^{-T/2}\big|\mathbf{V}^{-1}\big|^{T/2}e^{-\frac{1}{2}(\mathrm{vec}([\mathbf{Y},\mathbf{Z}])-\mathbf{X}_{\mathrm{SUR}}\boldsymbol{\beta})'(\mathbf{V}^{-1}\otimes\mathbf{I}_T)(\mathrm{vec}([\mathbf{Y},\mathbf{Z}])-\mathbf{X}_{\mathrm{SUR}}\boldsymbol{\beta})}, \tag{A.12b}$$

$$p(\boldsymbol{\beta}|\boldsymbol{\theta}) = (2\pi)^{-n_\beta/2}\big|\lambda_B\mathbf{X}'_{0,\mathrm{SUR}}(\boldsymbol{\theta})\big(\bar{\mathbf{V}}_0(\boldsymbol{\theta})^{-1}\otimes\mathbf{I}_{m(mp+k)}\big)\mathbf{X}_{0,\mathrm{SUR}}\big|^{1/2}$$

$$\times\, e^{-\frac{\lambda_B}{2}(\mathbf{X}_{0,\mathrm{SUR}}(\bar{\boldsymbol{\beta}}_0(\boldsymbol{\theta})-\boldsymbol{\beta}))'(\bar{\mathbf{V}}_0^{-1}\otimes\mathbf{I}_{m(mp+k)})(\mathbf{X}_{0,\mathrm{SUR}}(\bar{\boldsymbol{\beta}}_0(\boldsymbol{\theta})-\boldsymbol{\beta}))},$$

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

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

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

$$p(\boldsymbol{\theta}) = \mathbf{1}\{\text{DSGE model has a unique \& stable solution}|\boldsymbol{\theta}\}$$

$$\times\prod_{n=1}^{n_\theta}p_n\big(\boldsymbol{\theta}^{(n)}\big), \tag{A.12e}$$

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

Implementing the DSGE-SVAR prior The following dummy observations and likelihood implement the prior that $\boldsymbol{\beta}\sim\mathcal{N}(\bar{\boldsymbol{\beta}}_0,\mathbf{N}_{XX}(\bar{\mathbf{V}}_0))$ and $\mathbf{V}^{-1}\sim\mathcal{W}(\bar{\mathbf{V}}_0T_0^V,T_0^V)$:

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

$$\mathrm{vec}\big(\big[\mathbf{Y}_0^B,\mathbf{Z}_0^B\big]\big) \sim \mathcal{N}\big(\bar{\mathbf{X}}_{0,\mathrm{SUR}}(\boldsymbol{\theta})\bar{\boldsymbol{\beta}}_0(\boldsymbol{\theta}),\bar{\mathbf{V}}_0(\boldsymbol{\theta})\otimes\mathbf{I}_{T_0^B}\big), \tag{A.13b}$$

$$\big[\mathbf{Y}_0^V,\mathbf{Z}_0^V\big] = \mathbf{0}\times\boldsymbol{\beta} + \bar{\mathbf{V}}_0(\boldsymbol{\theta})\otimes\mathbf{I}_{T_0^V}, \tag{A.13c}$$

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

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

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

The prior and data density are given by

$$\boldsymbol{\theta}\sim\pi(\boldsymbol{\theta}), \tag{A.14a}$$

$$\pi\big(\mathbf{B},\mathbf{V}^{-1}|\boldsymbol{\theta}\big) \propto \big|\mathbf{V}^{-1}\big|^{-n_y/2}\ell\big(\mathbf{B},\mathbf{V}^{-1}|\mathbf{Y}_0(\boldsymbol{\theta}),\mathbf{Z}_0(\boldsymbol{\theta})\big)$$

$$= \big|\mathbf{V}^{-1}\big|^{-n_y/2}f\big(\mathbf{Y}_0(\boldsymbol{\theta}),\mathbf{Z}_0(\boldsymbol{\theta})|\mathbf{B},\mathbf{V}^{-1}\big), \tag{A.14b}$$

$$\tilde{f}\big(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{V}^{-1},\boldsymbol{\theta}\big) = f\big(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{V}^{-1}\big). \tag{A.14c}$$

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, Schorfheide, Smets, and Wouters (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 *i.i.d.* sample of length T with $y_t \in \mathbb{R}^m$ is given by

$$
\begin{aligned}
p(y|T_0^V) &\equiv \int_0^\infty \left(\prod_{s=1}^T f(y_s|V)\right)\pi(V|T_0^V)\,dV^{-1} \\
&= \int_0^\infty (2\pi)^{-mT/2}|\mathbf{V}|^{-T/2}e^{-\frac{T}{2}\operatorname{tr}(V^{-1}\hat{V})-\frac{T_0^V}{2}\operatorname{tr}(V^{-1}V_0)}2^{-mT_0^V/2} \\
&\quad \times |V_0T_0^V|^{T_0^V/2}\Gamma_m(T_0^V/2)^{-1}|\mathbf{V}|^{-(T_0-m-1)/2}\,dV^{-1} \\
&= \pi^{-mT/2}\frac{|V_0T_0^V|^{T_0^V/2}\Gamma_m((T+T_0^V)/2)}{|V_0T_0^V+T\hat{V}|^{(T+T_0^V)/2}\Gamma_m(T_0^V/2)}\times \\
&\quad \times \int_0^\infty e^{-\frac{1}{2}\operatorname{tr}(V^{-1}(T\hat{V}+T_0^VV_0))}2^{-m(T+T_0^V)/2} \\
&\quad \times |V_0T_0^V+T\hat{V}|^{(T+T_0^V)/2}\Gamma_m((T+T_0^V)/2)^{-1}|\mathbf{V}|^{-(T+T_0-2)/2}\,dV^{-1} \\
&= \pi^{-mT/2}\frac{|V_0T_0^V|^{T_0^V/2}\Gamma_m((T+T_0^V)/2)}{|V_0T_0^V+T\hat{V}|^{(T+T_0^V)/2}\Gamma_m(T_0^V/2)} \\
&= \pi^{-mT/2}|\hat{V}|^{-T/2}\frac{|\hat{V}^{-1}V_0|^{T_0^V/2}}{\left|\hat{V}^{-1}V_0+I_m\frac{T}{T_0^V}\right|^{(T+T_0^V)/2}}\frac{(T_0^V)^{-mT/2}\Gamma_m((T+T_0^V)/2)}{\Gamma_m(T_0^V/2)}
\end{aligned}
\tag{A.15}
$$

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

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

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

$$\frac{d\ln p(y|T_0^V)}{dT_0^V} = \frac{1}{2}\ln\left(\frac{T_0^V\frac{V_0}{\hat{V}}}{T_0^V\frac{V_0}{\hat{V}}+T}\right) + \frac{1}{2}\left(1-\frac{V_0}{\hat{V}}\right)\frac{T}{T_0^V\frac{V_0}{\hat{V}}+T}$$
$$+\frac{1}{2}\psi\left(\frac{T+T_0^V}{2}\right) - \frac{1}{2}\psi\left(\frac{T_0^V}{2}\right), \tag{A.16}$$

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

(a) *For $\frac{V_0}{\hat{V}}$ 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 $\frac{V_0}{\hat{V}} < \underline{v}$ $\frac{d}{dT_0^V}\ln p(y|T_0^V) < 0$.*

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

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.16) can therefore be written as

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

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: $-\log(\frac{x}{2+x})=\log(1+\frac{2}{x})<\frac{2}{x}$. Thus, $\frac{1}{2}\ln(\frac{x}{2+x})+\frac{1}{x}>0$.

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

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

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

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 $\lim_{V_0/\hat{V}\to\infty}\frac{d}{dT_0^V}\ln p(y|T_0^V)=-\frac{T}{2T_0^V}+\frac{1}{2}\psi(\frac{T+T_0^V}{2})-\frac{1}{2}\psi(\frac{T_0^V}{2})$. Note also that $\psi(\frac{T+T_0^V}{2})-\psi(\frac{T_0^V}{2})\leq\frac{T}{2}\frac{2}{T_0}$ given the recurrence relation used in (a) and given that the sum in the recurrence 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.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 $\bar{\mathbf{B}}_0^y$, $\bar{\mathbf{B}}_0^z$ and $\bar{\mathbf{V}}_0$ matrices so that the prior fit is perfect: I choose the prior to equal the posterior

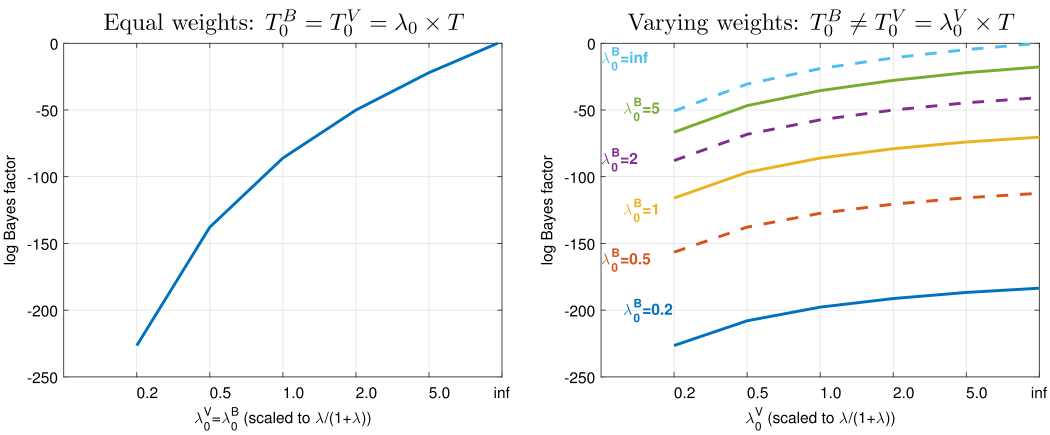

FIGURE A.1. Narrative DSGE-VAR marginal likelihood with fixed hyperparameters when prior is set to equal the posterior. In this numerical example, the prior is chosen to equal the posterior for the baseline narrative DSGE-VAR. Thus the prior fits the data as well as possible. The figures show that increasing 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.

given the actual data. I then vary the prior precision T_0^B and T_0^V on a grid. Figure A.1 shows the results. 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.

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

$$\pi(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) = \frac{p(\mathbf{y}, \mathbf{z}|\mathbf{V}^{-1}, \mathbf{B})\pi(\mathbf{V}^{-1}, \mathbf{B}|\boldsymbol{\theta})}{\pi(\hat{\mathbf{z}}, \mathbf{V}^{-1}, \mathbf{B}|\mathbf{y}, \mathbf{z}, \theta)} = \frac{p(\mathbf{y}, \mathbf{z}|\mathbf{V}^{-1}, \mathbf{B})\pi(\mathbf{V}^{-1}, \mathbf{B}|\boldsymbol{\theta})}{\pi(\mathbf{B}_*|\mathbf{y}, \mathbf{z}, \boldsymbol{\theta})\pi(\mathbf{V}_*^{-1}|\mathbf{B}_*, \mathbf{y}, \mathbf{Z}, \boldsymbol{\theta})} \tag{A.17}$$

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

$$\pi(\mathbf{B}_*|\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = M^{-1} \sum_{m=1}^{M} \pi\big(\mathbf{B}_*|\mathbf{y}, \mathbf{z}, \big(\mathbf{V}^{-1}\big)^{(m)}, \hat{\mathbf{z}}^{(m)}, \boldsymbol{\theta}\big),$$

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

$$\pi\big(\mathbf{V}_*^{-1}|\mathbf{y},\mathbf{z},\boldsymbol{\theta}\big) = M^{-1}\sum_{m=1}^{M}\pi\big(\mathbf{V}_*^{-1}|\mathbf{y},\mathbf{z},\mathbf{B}_*,\hat{\mathbf{z}}^{(m)},\boldsymbol{\theta}\big),$$

where $(\mathbf{V}^{-1})^{(m)}$, $\hat{\mathbf{z}}^{(m)}$ are draws from a simpler new run of the Gibbs sampler that conditions on $\mathbf{B}_*$.

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

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

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

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

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

where $\boldsymbol{\psi}_m$ are draws from the posterior.

Here, $\boldsymbol{\psi} = (\boldsymbol{\theta},\mathbf{B},\mathbf{V}^{-1})$—or strictly $(\boldsymbol{\theta},\mathbf{B},\mathrm{vech}(\mathbf{V}^{-1}))$. This vector is high dimensional, especially because of the presence of $\mathbf{B}$. It would therefore be helpful to reduce the dimensionality of the parameter vector, which I do using the Chib (1995) algorithm previously described:

$$\begin{aligned}
& k\big(\boldsymbol{\theta},\mathbf{B},\mathbf{V}^{-1}\big) = p\big(\mathbf{y},\mathbf{z}|\mathbf{B},\mathbf{V}^{-1}\big)p\big(\mathbf{B},\mathbf{V}^{-1}|\boldsymbol{\theta}\big)\pi(\boldsymbol{\theta}) \\
\Rightarrow\quad & k(\boldsymbol{\theta}) \equiv \int\!\!\int k\big(\boldsymbol{\theta},\mathbf{B},\mathbf{V}^{-1}\big)\,d\mathbf{V}^{-1}\,d\mathbf{B} \\
& = \pi(\boldsymbol{\theta})\int\!\!\int p\big(\mathbf{y},\mathbf{z}|\mathbf{B},\mathbf{V}^{-1}\big)p\big(\mathbf{B},\mathbf{V}^{-1}|\boldsymbol{\theta}\big)\,d\mathbf{V}^{-1}\,d\mathbf{B} \\
& \equiv \pi(\boldsymbol{\theta})\int\!\!\int k\big(\mathbf{B},\mathbf{V}^{-1}|\mathbf{y},\mathbf{z},\theta\big)\,d\mathbf{V}^{-1}\,d\mathbf{B} \\
& = \pi(\boldsymbol{\theta})p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})\int\!\!\int p\big(\mathbf{B},\mathbf{V}^{-1}|\mathbf{y},\mathbf{z},\boldsymbol{\theta}\big)\,d\mathbf{V}^{-1}\,d\mathbf{B} \\
\Leftrightarrow\quad & k(\boldsymbol{\theta}) = p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})\pi(\boldsymbol{\theta}).
\end{aligned}$$

Now proceed with this reduced parameter vector as before

$$\begin{aligned}\mathbb{E}[g(\boldsymbol{\theta})] &= \int_\Theta \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{z})\,d\boldsymbol{\theta} = \int_\Theta \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{z})\,d\boldsymbol{\theta} \\ &= p(\mathbf{y},\mathbf{z})^{-1}\int_\Theta \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})}\pi(\boldsymbol{\theta})p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})\,d\boldsymbol{\theta} \\ &= p(\mathbf{y},\mathbf{z})^{-1}\int_\Theta f(\boldsymbol{\theta})\,d\boldsymbol{\theta} = p(\mathbf{y},\mathbf{z})^{-1}.\end{aligned}$$

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

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

where $\hat{p}(\mathbf{y},\mathbf{z},\boldsymbol{\theta}_{(m)})$ is the Chib estimator of the (conditional) marginal likelihood. The approximation relies on $\int_\Theta f(\boldsymbol{\theta})\frac{p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})}{\hat{p}(\mathbf{y},\mathbf{z},\boldsymbol{\theta})}\,d\boldsymbol{\theta}$ 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:

$$\hat{k}_t^p = \left(1 - \frac{\bar{x}}{\bar{k}^p}\right)\hat{k}_{t-1}^p + \frac{\bar{x}}{\bar{k}^p}(\hat{x}_t + \hat{q}_{t+s}). \tag{A.18}$$

Household wage setting:

$$\begin{aligned}\hat{w}_t &= \frac{\hat{w}_{t-1}}{1+\bar{\beta}\gamma} + \frac{\bar{\beta}\gamma\mathbb{E}_t[\hat{w}_{t+1}]}{1+\bar{\beta}\gamma} \\ &+ \frac{(1-\bar{\beta}\zeta_w\gamma)(1-\zeta_w)}{(1+\bar{\beta}\gamma)\zeta_w}\bar{A}_w\left(\frac{\hat{c}_t - (h/\gamma)\hat{c}_{t-1}}{1-h/\gamma} + \nu\hat{n}_t - \hat{w}_t + \frac{d\tau_t^n}{1-\bar{\tau}^n} + \frac{d\tau_t^c}{1+\bar{\tau}^c}\right) \\ &- \frac{1+\bar{\beta}\mu\iota_w}{1+\bar{\beta}\gamma}\hat{\pi}_t + \frac{\iota_w}{1+\bar{\beta}\gamma}\hat{\pi}_{t-1}\frac{\bar{\beta}\gamma}{1+\bar{\beta}\gamma}\mathbb{E}_t[\hat{\pi}_{t+1}] + \frac{\hat{\epsilon}_t^{\lambda,w}}{1+\bar{\beta}\gamma}.\end{aligned} \tag{A.19}$$

Household consumption Euler equation:

$$\begin{aligned}&\mathbb{E}_t[\hat{\xi}_{t+1} - \hat{\xi}_t] + \mathbb{E}_t\left[d\tau_{t+1}^c - d\tau_t^c\right] \\ &\quad = \frac{1}{1-h/\gamma}\mathbb{E}_t\Bigg((\sigma-1)\frac{1}{1+\bar{\lambda}_w}\frac{1-\bar{\tau}^n}{1+\tau^c}\frac{\bar{w}\bar{n}}{\bar{c}}[\hat{n}_{t+1}-\hat{n}_t] \\ &\quad\quad - \sigma\left[\hat{c}_{t+1} - \left(1+\frac{h}{\gamma}\right)c_t + \frac{h}{\gamma}\hat{c}_{t+1}\right]\Bigg).\end{aligned} \tag{A.20}$$

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

$$\mathbb{E}_t[\hat{\xi}_{t+1} - \hat{\xi}_t] = -\hat{q}_t^b - \hat{R}_t + \mathbb{E}_t[\hat{\pi}_{t+1}], \tag{A.21}$$

$$\begin{aligned}\hat{Q}_t = -\hat{q}_t^b - (\hat{R}_t - \mathbb{E}_t[\pi_{t+1}]) + \frac{1}{\bar{r}^k(1-\tau^k) + \delta\tau^k + 1 - \delta} \\ \times \big[(\bar{r}^k(1-\tau^k) + \delta\tau^k)\hat{q}_t^k - (\bar{r}^k - \delta)d\tau_{t+1}^k \qquad \text{(A.22)} \\ + \bar{r}^k(1-\tau^k)\mathbb{E}_t(\hat{r}_{t+1}^k) + (1-\delta)\mathbb{E}_t(\hat{Q}_{t+1})\big], \qquad \text{(A.23)}\end{aligned}$$

$$\hat{x}_t = \frac{1}{1+\bar{\beta}\gamma}\left[\hat{x}_{t-1} + \bar{\beta}\gamma\mathbb{E}_t(\hat{x}_{t+1}) + \frac{1}{\gamma^2 S''(\gamma)}[\hat{Q}_t + \hat{q}_t^x]\right], \tag{A.24}$$

$$\hat{u}_t = \frac{a'(1)}{a''(1)}\hat{r}_t^k \equiv \frac{1-\psi_u}{\psi_u}\hat{r}_t^k. \tag{A.25}$$

A.8.2 *Production side and price setting* The linearized aggregate production function is

$$\hat{y}_t = \frac{\bar{y} + \Phi}{\bar{y}}\big(\hat{\epsilon}_t^a + \zeta\hat{k}_{t-1}^g + \alpha(1-\zeta)\hat{k}_t + (1-\alpha)(1-\zeta)\hat{n}_t\big), \tag{A.26}$$

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

The capital-labor ratio:

$$\hat{k}_t = \hat{n}_t + \hat{w}_t - \hat{r}_t^k. \tag{A.27}$$

Price setting:

$$\hat{\pi}_t = \frac{\iota_p}{1+\iota_p\bar{\beta}\gamma}\hat{\pi}_{t-1} + \frac{1-\zeta_p\bar{\beta}\gamma}{1+\iota_p\bar{\beta}\gamma}\frac{1-\zeta_p}{\zeta_p}\bar{A}_p\big(\widehat{mc}_t + \hat{\epsilon}_t^{\lambda,p}\big) + \frac{\bar{\beta}\gamma}{1+\iota_p\bar{\beta}\gamma}\mathbb{E}_t\hat{\pi}_{t+1}. \tag{A.28}$$

Marginal costs with a cost-channel:

$$\widehat{mc}_t = \alpha\hat{r}_t^k + (1-\alpha)(\hat{w}_t + \hat{R}_t). \tag{A.29}$$

A.8.3 *Market clearing* Goods market clearing requires

$$\hat{y}_t = \frac{\bar{c}}{\bar{y}}\hat{c}_t + \frac{\bar{x}}{\bar{y}}\hat{x}_t + \frac{\bar{x}^g}{\bar{y}}\hat{x}_t^g + \hat{g}_t + \frac{\bar{r}^k\bar{k}}{\bar{y}}\hat{u}_t. \tag{A.30}$$

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:

$$\Delta \ln g_t^{\text{obs}} = g_t - g_{t+1} + (\gamma_g - 1), \tag{A.31a}$$

$$\Delta \ln x_t^{\text{obs}} = x_t - x_{t+1} + (\gamma_x - 1), \tag{A.31b}$$

$$\Delta \ln w_t^{\text{obs}} = w_t - w_{t+1} + (\gamma_w - 1), \tag{A.31c}$$

$$\Delta \ln c_t^{\text{obs}} = c_t - c_{t+1} + (\gamma - 1), \tag{A.31d}$$

$$\hat{\pi}_t^{\text{obs}} = \hat{\pi}_t + \bar{\pi}, \tag{A.31e}$$

$$\hat{n}_t^{\text{obs}} = \hat{n}_t + \bar{n}, \tag{A.31f}$$

$$\hat{R}_t^{\text{obs}} = \hat{R}_t + \left(\beta^{-1} - 1\right). \tag{A.31g}$$

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:

$$b_t^{\text{obs}} = \frac{\bar{b}}{\bar{y}}(\hat{b} - \hat{y}) + \bar{b}^{\text{obs}}, \tag{A.32a}$$

$$\text{rev}_t^{n,\text{obs}} = \bar{\tau}^n \frac{\bar{w}\bar{n}}{\bar{c}} \frac{\bar{c}}{\bar{y}} \left(\frac{d\tau_t^n}{\bar{\tau}^n} + \hat{w}_t + \hat{n}_t - \hat{y}_t \right) + \bar{\text{rev}}^{n,\text{obs}}, \tag{A.32b}$$

$$\text{rev}_t^{k,\text{obs}} = \bar{\tau}^k \frac{\bar{k}}{\bar{y}} (\bar{r}^k - \delta) \left(\frac{d\tau^k}{\bar{\tau}^k} + \frac{\bar{r}^k}{\bar{r}^k - \delta} \hat{r}_t^k + \hat{k}_{t-1}^p - \hat{y}_t \right) + \bar{\text{rev}}^{k,\text{obs}}. \tag{A.32c}$$

Appendix 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,

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

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

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

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

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

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

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

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

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

where ω 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

$$n_t^{FR} = \frac{(\text{Total hours worked: Francis and Ramey (2009)})_t}{(\text{Population above 16: FRED CNP16OV})_t}.$$

Fiscal data is computed following Leeper, Plante, and Traum (2010), except for adding state and local governments (superscript "s&l") to the federal government account (superscript "f"), similar to Fernandez-Villaverde, Guerron-Quintana, Kuester, and Rubio-Ramirez (2015). Since in the real world

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

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

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

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

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.
- (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.

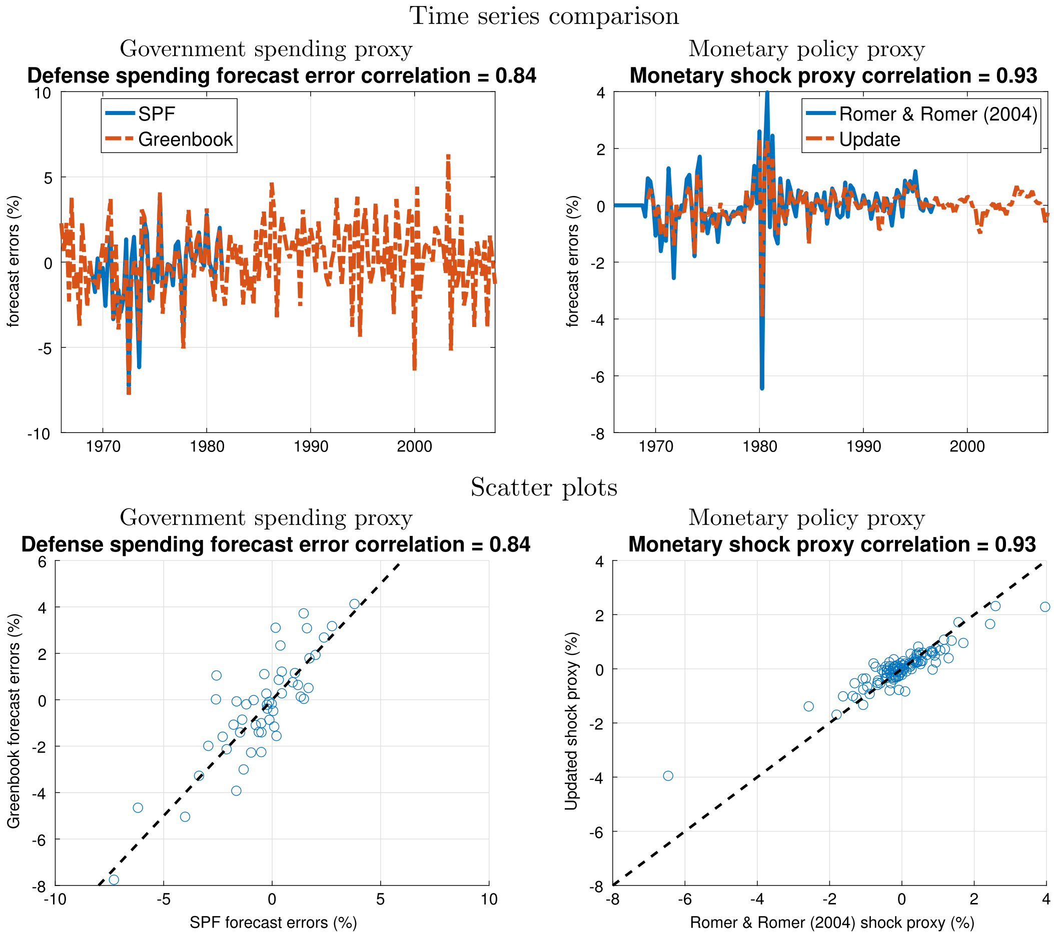

FIGURE B.1. Comparing shock proxies in the literature with their updated counterparts.

- 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.

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

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

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

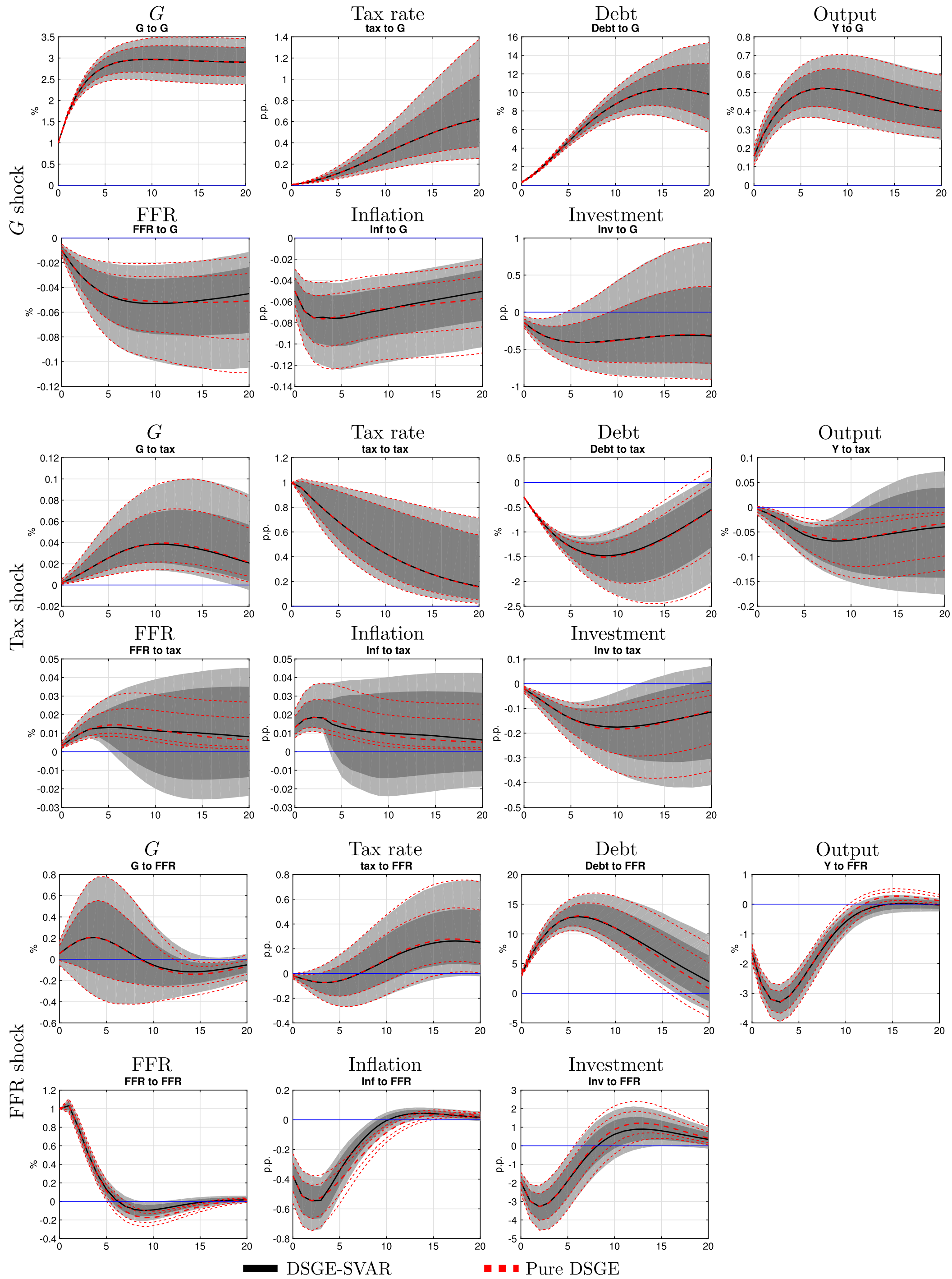

FIGURE B.2. Responses of output, investment and inflation: Quality of VAR approximation to DSGE model ($T_0^V = T_0^B \nearrow \infty$). *Note*: Shown are the pointwise median and 68% and 90% posterior credible sets. Results based on lower Cholesky factorization of $S_1 S_1'$.

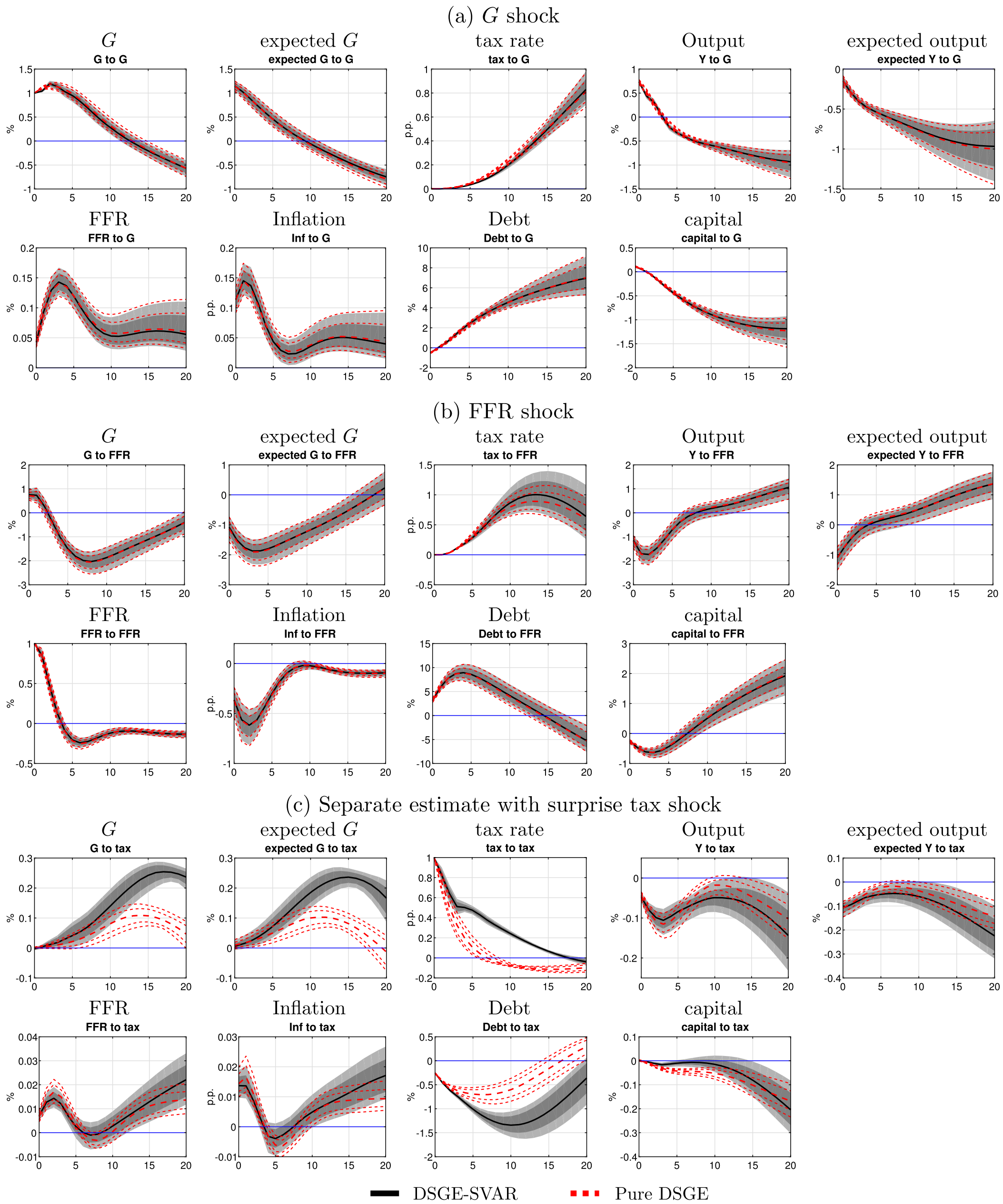

FIGURE B.3. Responses in expectations-augmented DSGE-VAR: Quality of VAR approximation.

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 expen-

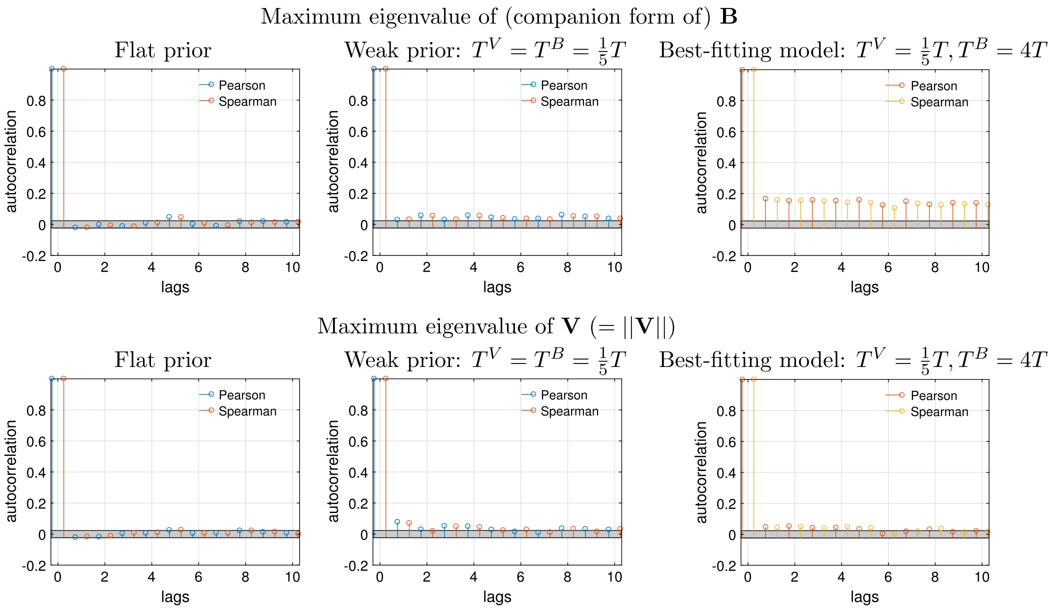

FIGURE B.4. Gibbs sampler of baseline model: Autocorrelation functions of univariate summary statistics by DSGE prior weight. *Note*: Autocorrelations are reported based on both the Pearson and the Spearman correlation measure. Asymptotic classical 90% credible sets for the Pearson coefficient, computed 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.

ditures 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.

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

[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).

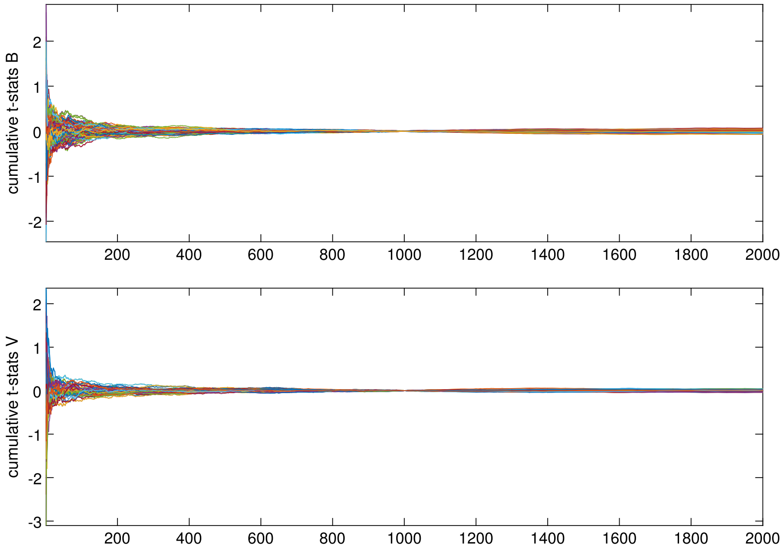

FIGURE B.5. Brooks and Gelman (1998) type convergence diagnostic for the flat-prior narrative VAR. Shown are the (within-chain) means of the parameter estimates as the Markov chain grows. To standardize the plots, the parameter estimates are displayed minus their mean and standard deviation in the first half of the chain: For example, for element i of $\boldsymbol{\theta}$, the plot shows $\frac{t^{-1}\sum_{s=1}^{t}\theta_s(i)-\lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}\theta_s(i)}{(\lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}(\theta_s(i)-\lfloor T/2\rfloor^{-1}\sum_{u=1}^{\lfloor T/2\rfloor}\theta_u(i))^2)^{1/2}}$ 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 very good for both the elements of the VAR coefficient matrix **B** and the covariance matrix **V**.

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

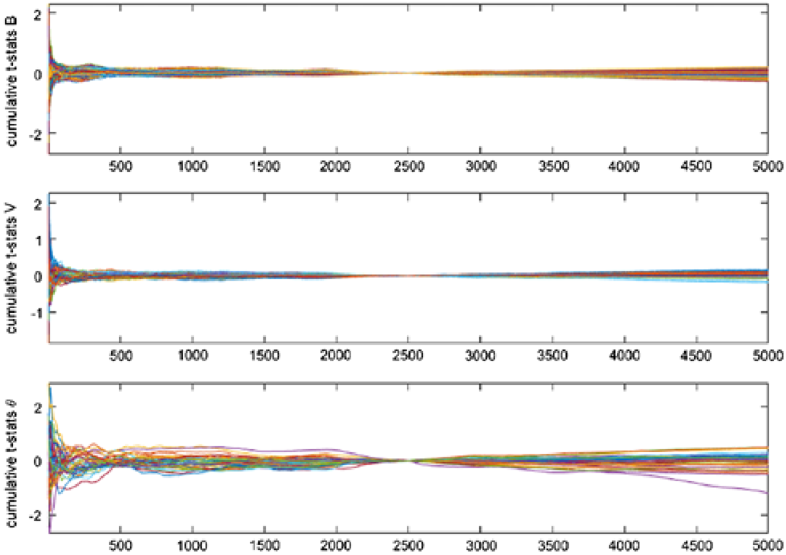

FIGURE B.6. Brooks and Gelman (1998) type convergence diagnostic for DSGE-VAR with best–fitting model ($T_0^B = 4T$, $T_0^V = \frac{1}{5}T$). Shown are the (within-chain) means of the parameter estimates as the Markov chain grows. To standardize the plots, the parameter estimates are displayed minus their mean and standard deviation in the first half of the chain: For example, for element i of $\boldsymbol{\theta}$, the plot shows $\frac{t^{-1}\sum_{s=1}^{t}\theta_s(i) - \lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}\theta_s(i)}{(\lfloor T/2\rfloor^{-1}\sum_{s=1}^{\lfloor T/2\rfloor}(\theta_s(i) - \lfloor T/2\rfloor^{-1}\sum_{u=1}^{\lfloor T/2\rfloor}\theta_u(i))^2)^{1/2}}$ as a function of t. Brooks and Gelman (1998) argued 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 4000 draws.

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.1 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: $\mathbb{E}_t^{GB}[g^n_{\text{Def},t+1} - \pi_{t+1}] - (g^n_{\text{Def},t} - \pi_t)$. The left panels in Figure B.1 compare my updated series with the SPF series. The correlation is 0.84 over the entire sample period with observed shocks.

B.2 *Approximation quality of VAR representation of DSGE model*

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.5 and B.6 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.4 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 5000 for the DSGE-VAR and 2000 for the flat prior VAR. This produces results consistent with convergence of the sampler (see Figures B.5 and B.6). The resulting samples are also reasonably efficient: the autocorrelation of the subsamples in Figure B.4 are reasonably small, particular with low prior weights on the DSGE model.

References

Brooks, S. P. and A. Gelman (1998), "General methods for monitoring convergence of iterative simulations." *Journal of Computational and Graphical Statistics*, 7, 434–455. [24, 25]

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2005), "A critique of structural VARs using real business cycle theory." Working Paper 631, Federal Reserve Bank of Minneapolis. [9]

Chib, S. (1995), "Marginal likelihood from the Gibbs output." *Journal of the American Statistical Association*, 90, 1313–1321. [14, 15, 16]

Chib, S. and S. Ramamurthy (2010), "Tailored randomized block MCMC methods with application to DSGE models." *Journal of Econometrics*, 155 (1), 19–38. [11]

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007), "On the fit of new Keynesian models." *Journal of Business & Economic Statistics*, 25, 123–143. [11]

Fernandez-Villaverde, J., P. A. Guerron-Quintana, K. Kuester, and J. Rubio-Ramirez (2015), "Fiscal volatility shocks and economic activity." *American Economic Review*, 105 (11), 3352–3384. [19]

Fernandez-Villaverde, J., J. F. Rubio-Ramirez, T. J. Sargent, and M. W. Watson (2007), "ABCs (and Ds) of understanding VARs." *American Economic Review*, 97 (3), 1021–1026. [8, 9]

Francis, N. and V. A. Ramey (2009), "Measures of per capita hours and their implications for the technology-hours debate." *Journal of Money, Credit and Banking*, 41 (6), 1071–1097. [19]

Geweke, J. (1999), "Using simulation methods for Bayesian econometric models: Inference, development, and communication." *Econometric Reviews*, 18 (1), 1–73. [14, 15]

Leeper, E. M., M. Plante, and N. Traum (2010), "Dynamics of fiscal financing in the United States." *Journal of Econometrics*, 156 (2), 304–321. [19]

Mertens, K. and M. O. Ravn (2013), "The dynamic effects of personal and corporate income tax changes in the United States." *American Economic Review*, 103 (4), 1212–1247. [3, 4, 24]

Ramey, V. A. (2011), “Identifying government spending shocks: It’s all in the timing.” *Quarterly Journal of Economics*, 126 (1), 1–50. [24, 25]

Romer, C. D. and D. H. Romer (2004), “A new measure of monetary shocks: Derivation and implications.” *American Economic Review*, 94 (4), 1055–1084. [24, 25]

Smets, F. and R. Wouters (2007), “Shocks and frictions in US business cycles: A Bayesian DSGE approach.” *American Economic Review*, 97 (3), 586–606. [11, 17, 18]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 1 March, 2018; final version accepted 2 June, 2019; available online 11 December, 2019.