Supplement to "Behavioral learning equilibria in New Keynesian models"

(*Quantitative Economics*, Vol. 14, No. 4, November 2023, 1401–1445)

Cars Hommes
Bank of Canada and CeNDEF, Amsterdam School of Economics, University of Amsterdam

Kostas Mavromatis
De Nederlandsche Bank and Amsterdam School of Economics, University of Amsterdam

Tolga Özden
Bank of Canada

Mei Zhu
Institute for Advanced Research and School of Economics, Shanghai University of Finance and Economics and The Key Laboratory of Mathematical Economics (SUFE), Ministry of Education

Appendix A: Theoretical results

A.1 *Mean of the rational expectations equilibrium*

Using (2.10)–(2.11) and (2.15)–(2.18), the mean of the REE satisfies

$$\begin{aligned}
\overline{\mathbf{x}^*} &= (\boldsymbol{I} - \boldsymbol{c}_1)^{-1}(\boldsymbol{c}_0 + \boldsymbol{c}_2\overline{\boldsymbol{u}}) \\
&= (\boldsymbol{I} - \boldsymbol{c}_1)^{-1}(\boldsymbol{I} - \boldsymbol{b}_1\boldsymbol{c}_1 - \boldsymbol{b}_1)^{-1}(\boldsymbol{b}_0 + \boldsymbol{b}_1\boldsymbol{c}_2\boldsymbol{a}) + (\boldsymbol{I} - \boldsymbol{c}_1)^{-1}\boldsymbol{c}_2(\boldsymbol{I} - \boldsymbol{\rho})^{-1}\boldsymbol{a} \\
&= (\boldsymbol{I} - \boldsymbol{c}_1)^{-1}(\boldsymbol{I} - \boldsymbol{b}_1\boldsymbol{c}_1 - \boldsymbol{b}_1)^{-1}\big[\boldsymbol{b}_0 + \big(\boldsymbol{b}_1\boldsymbol{c}_2(\boldsymbol{I} - \boldsymbol{\rho}) + (\boldsymbol{I} - \boldsymbol{b}_1\boldsymbol{c}_1 - \boldsymbol{b}_1)\boldsymbol{c}_2\big)(\boldsymbol{I} - \boldsymbol{\rho})^{-1}\boldsymbol{a}\big] \\
&= \big[(\boldsymbol{I} - \boldsymbol{b}_1\boldsymbol{c}_1 - \boldsymbol{b}_1)(\boldsymbol{I} - \boldsymbol{c}_1)\big]^{-1}\big[\boldsymbol{b}_0 + \boldsymbol{b}_3(\boldsymbol{I} - \boldsymbol{\rho})^{-1}\boldsymbol{a}\big] \\
&= (\boldsymbol{I} - \boldsymbol{b}_1 - \boldsymbol{b}_2)^{-1}\big[\boldsymbol{b}_0 + \boldsymbol{b}_3(\boldsymbol{I} - \boldsymbol{\rho})^{-1}\boldsymbol{a}\big].
\end{aligned}$$

A.2 *Autocorrelations in the multivariate linear model*

The purpose of this Appendix is to compute the first-order autocorrelation coefficients of the linear stochastic stationary system (2.24) and to show that these are continuous functions with respect to $(\beta_1, \beta_2, \dots, \beta_n)$ and the other parameters.

Cars Hommes: C.H.Hommes@uva.nl
Kostas Mavromatis: K.Mavromatis@dnb.nl
Tolga Özden: tozden@bank-banque-canada.ca
Mei Zhu: Zhu.Mei@mail.shufe.edu.cn
The views expressed in this paper do not represent the positions of the Bank of Canada, De Nederlandsche Bank, or the Eurosystem.

 https://doi.org/10.3982/QE1533

Define $\boldsymbol{X}_t' = [\boldsymbol{x}_t', \boldsymbol{u}_t'] - [\overline{\boldsymbol{x}}', \overline{\boldsymbol{u}}']$. Rewrite model (2.24) as

$$\boldsymbol{X}_t = \widehat{\boldsymbol{B}}(\boldsymbol{\beta})\boldsymbol{X}_{t-1} + \widehat{\boldsymbol{C}}\boldsymbol{\eta}_t, \tag{A.1}$$

where $\boldsymbol{\eta}_t' = [\boldsymbol{v}_t', \boldsymbol{\varepsilon}_t']$, $\widehat{\boldsymbol{B}}(\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{b_1\beta}^2+\boldsymbol{b}_2 & \boldsymbol{b}_3\boldsymbol{\rho} \\ \mathbf{0} & \boldsymbol{\rho} \end{pmatrix}$, $\widehat{\boldsymbol{C}} = \begin{pmatrix} \boldsymbol{b}_4 & \boldsymbol{b}_3 \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$. The variance-covariance matrix $\widehat{\boldsymbol{\Gamma}}(0)$ and the autocovariance matrix $\widehat{\boldsymbol{\Gamma}}(1)$ satisfy

$$\widehat{\boldsymbol{\Gamma}}(0) = E\big[\boldsymbol{X}_t\boldsymbol{X}_t'\big] = \widehat{\boldsymbol{B}}(\boldsymbol{\beta})\widehat{\boldsymbol{\Gamma}}(0)\widehat{\boldsymbol{B}}'(\boldsymbol{\beta}) + \widehat{\boldsymbol{C}}\boldsymbol{\Sigma_\eta}\widehat{\boldsymbol{C}}', \tag{A.2}$$

$$\widehat{\boldsymbol{\Gamma}}(1) = E\big[\boldsymbol{X}_t\boldsymbol{X}_{t+1}'\big] = \widehat{\boldsymbol{\Gamma}}(0)\widehat{\boldsymbol{B}}'(\boldsymbol{\beta}), \tag{A.3}$$

where $\boldsymbol{\Sigma_\eta} = \begin{pmatrix} \boldsymbol{\Sigma}_v & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\varepsilon \end{pmatrix}$.

In order to obtain an expression for $\widehat{\boldsymbol{\Gamma}}(0)$, we use column stacks of matrices. Suppose $\mathrm{vec}(\boldsymbol{K})$ is the vectorization of a matrix $\boldsymbol{K}$ and $\otimes$ is the Kronecker product.[1] Under the assumption that all eigenvalues of $\boldsymbol{b_1\beta}^2 + \boldsymbol{b}_2$ and $\boldsymbol{\rho}$ are inside the unit circle, based on a property of Kronecker product,[2] it is easy to see that all eigenvalues of $\widehat{\boldsymbol{B}}(\boldsymbol{\beta}) \otimes \widehat{\boldsymbol{B}}(\boldsymbol{\beta})$ lie inside the unit circle, and hence $[\boldsymbol{I} - \widehat{\boldsymbol{B}}(\boldsymbol{\beta}) \otimes \widehat{\boldsymbol{B}}(\boldsymbol{\beta})]^{-1}$ exist. Therefore,

$$\mathrm{vec}\big(\widehat{\boldsymbol{\Gamma}}(0)\big) = \big[\boldsymbol{I} - \widehat{\boldsymbol{B}}(\boldsymbol{\beta}) \otimes \widehat{\boldsymbol{B}}(\boldsymbol{\beta})\big]^{-1}(\widehat{\boldsymbol{C}} \otimes \widehat{\boldsymbol{C}})\,\mathrm{vec}(\boldsymbol{\Sigma_\eta}). \tag{A.4}$$

From (A.4), we can obtain an expression for $\widehat{\boldsymbol{\Gamma}}(0)$, and using (A.3), an expression for $\widehat{\boldsymbol{\Gamma}}(1)$ can be obtained.

Based on the properties of matrix operations, it is easy to see that the entries of matrices $\widehat{\boldsymbol{\Gamma}}(0)$ and $\widehat{\boldsymbol{\Gamma}}(1)$ are continuous functions with respect to $(\beta_1, \beta_2, \dots, \beta_n)$ and the other parameters. Let $\boldsymbol{\Omega} = \mathrm{diag}[\gamma_{11}(0), \gamma_{22}(0), \dots, \gamma_{nn}(0)]$ (a diagonal matrix), where $\gamma_{ii}(0)(i = 1, \dots, n)$ are the first n diagonal entries of $\widehat{\boldsymbol{\Gamma}}(0)$. Let $\boldsymbol{L} = \mathrm{diag}[\gamma_{11}(1), \gamma_{22}(1), \dots, \gamma_{nn}(1)]$ (a diagonal matrix), where $\gamma_{ii}(1)$ $(i = 1, \dots, n)$ are the first n diagonal entries of $\widehat{\boldsymbol{\Gamma}}(1)$. Thus, the first-order autocorrelation coefficients of the linear stochastic stationary system (2.24) $\boldsymbol{G} = \boldsymbol{L\Omega}^{-1}$ are continuous functions with respect to $(\beta_1, \beta_2, \dots, \beta_n)$ and the other parameters.

For example, in the case $n = 1$, following the procedures above with Mathematica or Matlab software, one obtains

$$\begin{aligned} G(\beta) = \big[&\big(b_3^2\sigma_\varepsilon^2 + b_4^2\sigma_v^2\big)\big(b_1\beta^2 + b_2\big) + \big(b_3^2\sigma_\varepsilon^2 - b_4^2\sigma_v^2\big(b_1\beta^2 + b_2\big)^2\big)\rho \\ &- b_4^2\sigma_v^2\big(b_1\beta^2 + b_2\big)\rho^2 + b_4^2\sigma_v^2\big(b_1\beta^2 + b_2\big)^2\rho^3\big] \\ &/\big[b_3^2\sigma_\varepsilon^2\big(1 + \big(b_1\beta^2 + b_2\big)\rho\big) + b_4^2\sigma_v^2\big(1 - \big(b_1\beta^2 + b_2\big)\rho\big)\big(1 - \rho^2\big)\big]. \end{aligned}$$

In the special case $b_2 = 0$ and $b_4 = 1$, this expression is exactly the same as the first-order autocorrelation in Hommes and Zhu (2014), which was calculated using another approach.

[1] One property of column stacks is that the column stack of a product of three matrices is $\mathrm{vec}(ABC) = (C' \otimes A)\,\mathrm{vec}(B)$. For more details on this and related properties, see Magnus and Neudecker (2019, Chapter 2) and Evans and Honkapohja (2001, Section 5.7).

[2] The eigenvalues of $\check{A} \otimes \check{B}$ are the mn numbers $\lambda_r\mu_s$, $r = 1, 2, \dots, m$, $s = 1, 2, \dots, n$ where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of $m \times m$ matrix $\check{A}$ and $\mu_1, \dots, \mu_n$ are the eigenvalues of $n \times n$ matrix $\check{B}$ (see Lancaster and Tismenetsky (1985)).

A.3 *Proof of Proposition 2 (stability under SAC-learning)*

This Appendix derives the E-stability conditions for a BLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$. Set $\gamma_t = (1+t)^{-1}$. For the learning dynamics in (2.28) and (2.8),[3] since all functions are smooth, the SAC-learning rule satisfies the conditions (A.1)–(A.3) of Section 6.2.1 in Evans and Honkapohja (2001, p. 124). In order to check the conditions (B.1)–(B.2) of Section 6.2.1 in Evans and Honkapohja (2001, p. 125), we rewrite the system in matrix form as

$$\boldsymbol{X}_t = \widetilde{\boldsymbol{A}}(\boldsymbol{\theta}_{t-1})\boldsymbol{X}_{t-1} + \widetilde{\boldsymbol{B}}(\boldsymbol{\theta}_{t-1})\boldsymbol{W}_t,$$

where $\boldsymbol{\theta}'_t = (\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t, \boldsymbol{R}_t)$, $\boldsymbol{X}'_t = (1, \boldsymbol{x}'_t, \boldsymbol{x}'_{t-1}, \boldsymbol{u}'_t)$ and $\boldsymbol{W}'_t = (1, \boldsymbol{v}'_t, \boldsymbol{\varepsilon}'_t)$,

$$\widetilde{\boldsymbol{A}}(\boldsymbol{\theta}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \boldsymbol{b_0} + \boldsymbol{b_1}(\boldsymbol{I} - \boldsymbol{\beta}^2)\boldsymbol{\alpha} + \boldsymbol{b_3 a} & \boldsymbol{b_1}\boldsymbol{\beta}^2 + \boldsymbol{b_2} & \mathbf{0} & \boldsymbol{b_3\rho} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{a} & \mathbf{0} & \mathbf{0} & \boldsymbol{\rho} \end{pmatrix},$$

$$\widetilde{\boldsymbol{B}}(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 \\ \mathbf{0} & \boldsymbol{b_4} & \boldsymbol{b_3} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

Based on the properties of eigenvalues (see, e.g., Evans and Honkapohja (2001, p. 117)), all the eigenvalues of $\widetilde{\boldsymbol{A}}(\boldsymbol{\theta})$ include 0 (multiple $n+1$), the eigenvalues of $\boldsymbol{\rho}$ and $\boldsymbol{b_1}\boldsymbol{\beta}^2 + \boldsymbol{b_2}$. Thus, based on the assumptions, all the eigenvalues of $\widetilde{\boldsymbol{A}}(\boldsymbol{\theta})$ lie inside the unit circle. Moreover, it is easy to see all the other conditions in Section 6.2.1 in Evans and Honkapohja (2001) are also satisfied.

Since $\boldsymbol{x}_t$ is stationary, then the limits

$$\sigma_i^2 := \lim_{t\to\infty} E(x_{i,t} - \alpha_i)^2, \qquad \sigma^2_{x_i x_{i,-1}} := \lim_{t\to\infty} E(x_{i,t} - \alpha_i)(x_{i,t-1} - \alpha_i)$$

exist and are finite. Hence, according to Section 6.2.1 in Evans and Honkapohja (2001, p. 126), the associated ODE is

$$\begin{cases} \dfrac{d\boldsymbol{\alpha}}{d\tau} = \overline{\boldsymbol{x}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \boldsymbol{\alpha}, \\ \dfrac{d\boldsymbol{\beta}}{d\tau} = \boldsymbol{R}^{-1}[\boldsymbol{L} - \boldsymbol{\beta}\boldsymbol{\Omega}] = \boldsymbol{R}^{-1}\boldsymbol{\Omega}\big[\boldsymbol{L}\boldsymbol{\Omega}^{-1} - \boldsymbol{\beta}\big], \\ \dfrac{d\boldsymbol{R}}{d\tau} = \boldsymbol{\Omega} - \boldsymbol{R}, \end{cases} \tag{A.5}$$

where $\boldsymbol{R}$ is a diagonal matrix with the ith diagonal entry R_i and $\boldsymbol{\Omega}$, $\boldsymbol{L}$ are also diagonal matrices, as defined in Section 2. As shown in Evans and Honkapohja (2001), a BLE corresponds to a fixed point of the following ODE (A.6):

$$\begin{cases} \dfrac{d\boldsymbol{\alpha}}{d\tau} = \overline{\boldsymbol{x}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \boldsymbol{\alpha}, \\ \dfrac{d\boldsymbol{\beta}}{d\tau} = \boldsymbol{G} - \boldsymbol{\beta}. \end{cases} \tag{A.6}$$

[3] For convenience of theoretical analysis, one can set $\mathbf{S}_{t-1} = \mathbf{R}_t$.

Note that $\boldsymbol{\beta}$ and $\boldsymbol{G}$ are both diagonal matrices. The Jacobian matrix of (A.6) is, in fact, equivalent to

$$\begin{pmatrix} \left(\boldsymbol{I} - \boldsymbol{b}_1\boldsymbol{\beta}^{*2} - \boldsymbol{b}_2\right)^{-1}(\boldsymbol{b}_1 + \boldsymbol{b}_2 - \boldsymbol{I}) & \boldsymbol{\varrho} \\ \mathbf{0} & \boldsymbol{DG_\beta}(\boldsymbol{\beta}^*) - \boldsymbol{I} \end{pmatrix},$$

where $\boldsymbol{DG_\beta}$ is a Jacobian matrix with the (i, j)-th entry $\frac{\partial G_i}{\partial \beta_j}$, and the form of matrix $\boldsymbol{\varrho}$ is omitted since it is not needed in the proof. Therefore, if all the eigenvalues of $(\boldsymbol{I} - \boldsymbol{b}_1\boldsymbol{\beta}^{*2} - \boldsymbol{b}_2)^{-1}(\boldsymbol{b}_1 + \boldsymbol{b}_2 - \boldsymbol{I})$ have negative real parts and all the eigenvalues of $\boldsymbol{DG_\beta}(\boldsymbol{\beta}^*)$ have real parts less than 1, the SAC-learning $(\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t)$ converges to the BLE $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ as time t tends to ∞.

A.4 *Eigenvalues of matrix $\mathbf{B}\boldsymbol{\beta}^2$ inside the unit circle*

This Appendix shows the sufficiency condition for the existence of a BLE of (2.40) in Corollary 1. The characteristic polynomial of $\mathbf{B}\boldsymbol{\beta}^2$ is given by $h(\nu) = \nu^2 + c_1\nu + c_2$, where

$$c_1 = -\frac{\beta_y^2 + \left[\gamma\varphi + \lambda(1 + \varphi\phi_y)\right]\beta_\pi^2}{1 + \gamma\varphi\phi_\pi + \varphi\phi_y}, \qquad c_2 = \frac{\lambda\beta_y^2\beta_\pi^2}{1 + \gamma\varphi\phi_\pi + \varphi\phi_y}.$$

Both of the eigenvalues of $\mathbf{B}\boldsymbol{\beta}^2$ are inside the unit circle if and only if both of the following conditions hold (see Elaydi (2005)):

$$h(1) > 0, \qquad h(-1) > 0, \qquad \left|h(0)\right| < 1.$$

It is easy to see $h(-1) > 0$, $|h(0)| < 1$ for any $\beta_i \in [-1, 1]$. Note that

$$\begin{aligned} h(1) &= \frac{\left(1 - \beta_y^2\right)\left(1 - \lambda\beta_\pi^2\right) + \gamma\varphi\phi_\pi + \varphi\phi_y - (\gamma\varphi + \lambda\varphi\phi_y)\beta_\pi^2}{1 + \gamma\varphi\phi_\pi + \varphi\phi_y} \\ &\geq \frac{\varphi\left[\gamma(\phi_\pi - 1) + (1 - \lambda)\phi_y\right]}{1 + \gamma\varphi\phi_\pi + \varphi\phi_y}. \end{aligned}$$

Thus, if $\gamma(\phi_\pi - 1) + (1 - \lambda)\phi_y > 0$, then $h(1) > 0$. Therefore, both eigenvalues of $\mathbf{B}\boldsymbol{\beta}^2$ lie inside the unit circle for all $\beta_i \in [-1, 1]$.

A.5 *First-order autocorrelations in the baseline NK model*

This Appendix gives the expressions for the first-order autocorrelation coefficients for the output gap and inflation in the NK baseline model. Through complicated calculations, the following expressions in terms of the structural parameters are obtained:

$$G_1(\beta_y, \beta_\pi) = \frac{\widetilde{f}_1}{\widetilde{g}_1}, \tag{A.7}$$

$$G_2(\beta_y, \beta_\pi) = \frac{\widetilde{f}_2}{\widetilde{g}_2}, \tag{A.8}$$

where

$$\begin{aligned}
\widetilde{f}_1 &= \sigma_y^2\big\{\big(\rho+\lambda_1+\lambda_2-\lambda\beta_\pi^2\big)\big[1-\lambda\beta_\pi^2(\rho+\lambda_1+\lambda_2)\big] \\
&\quad+\big[\lambda\beta_\pi^2(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)-\rho\lambda_1\lambda_2\big]\big[(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)-\lambda\beta_\pi^2\rho\lambda_1\lambda_2\big]\big\} \\
&\quad+\sigma_\pi^2\big\{\big(\varphi\phi_\pi(\rho+\lambda_1+\lambda_2)-\varphi\beta_\pi^2)\big)\big[\varphi\phi_\pi-\varphi\beta_\pi^2(\rho+\lambda_1+\lambda_2)\big] \\
&\quad+\big[\varphi\beta_\pi^2(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)-\varphi\phi_\pi\rho\lambda_1\lambda_2\big] \\
&\quad\cdot\big[\varphi\phi_\pi(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)-\varphi\beta_\pi^2\rho\lambda_1\lambda_2\big]\big\}, \\
\widetilde{g}_1 &= \sigma_y^2\big\{\big[\big(1+\lambda^2\beta_\pi^4\big)-2\lambda\beta_\pi^2(\rho+\lambda_1+\lambda_2)+\big(1+\lambda^2\beta_\pi^4\big)(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)\big] \\
&\quad-\rho\lambda_1\lambda_2\big[\big(1+\lambda^2\beta_\pi^4\big)(\rho+\lambda_1+\lambda_2)-2\lambda\beta_\pi^2(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2) \\
&\quad+\big(1+\lambda^2\beta_\pi^4\big)\rho\lambda_1\lambda_2\big]\big\}+\sigma_\pi^2\big\{\big[\big((\varphi\phi_\pi)^2+\varphi^2\beta_\pi^4\big)-2\varphi\phi_\pi\varphi\beta_\pi^2(\rho+\lambda_1+\lambda_2) \\
&\quad+\big((\varphi\phi_\pi)^2+\varphi^2\beta_\pi^4\big)(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)\big] \\
&\quad-\rho\lambda_1\lambda_2\big[\big((\varphi\phi_\pi)^2+\varphi^2\beta_\pi^4\big)(\rho+\lambda_1+\lambda_2)-2\varphi\phi_\pi\varphi\beta_\pi^2 \\
&\quad\cdot(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)+\big((\varphi\phi_\pi)^2+\varphi^2\beta_\pi^4\big)\rho\lambda_1\lambda_2\big]\big\}, \qquad \text{(A.9)} \\
\widetilde{f}_2 &= \sigma_y^2\big\{\gamma^2\big[(\rho+\lambda_1+\lambda_2)-\rho\lambda_1\lambda_2(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)\big]\big\} \\
&\quad+\sigma_\pi^2\big\{\big[(1+\varphi\phi_y)(\rho+\lambda_1+\lambda_2)-\beta_y^2\big]\cdot\big[(1+\varphi\phi_y)-\beta_y^2(\rho+\lambda_1+\lambda_2)\big] \\
&\quad+\big[\beta_y^2(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)-(1+\varphi\phi_y)\rho\lambda_1\lambda_2\big] \\
&\quad\cdot\big[(1+\varphi\phi_y)(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)-\beta_y^2\rho\lambda_1\lambda_2\big]\big\}, \\
\widetilde{g}_2 &= \sigma_y^2\big\{\gamma^2\big[1+\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2-\rho\lambda_1\lambda_2(\rho+\lambda_1+\lambda_2)-(\rho\lambda_1\lambda_2)^2\big]\big\} \\
&\quad+\sigma_\pi^2\big\{\big[\big((1+\varphi\phi_y)^2+\beta_y^4\big)-2(1+\varphi\phi_y)\beta_y^2(\rho+\lambda_1+\lambda_2)+\big((1+\varphi\phi_y)^2+\beta_y^4\big) \\
&\quad\cdot(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)\big]-\rho\lambda_1\lambda_2\big[\big((1+\varphi\phi_y)^2+\beta_y^4\big)(\rho+\lambda_1+\lambda_2) \\
&\quad-2(1+\varphi\phi_y)\beta_y^2 \\
&\quad\cdot(\rho\lambda_1+\rho\lambda_2+\lambda_1\lambda_2)+\big((1+\varphi\phi_y)^2+\beta_y^4\big)\rho\lambda_1\lambda_2\big]\big\}, \qquad \text{(A.10)}
\end{aligned}$$

$$\lambda_1+\lambda_2=\frac{\beta_y^2+(\gamma\varphi+\lambda+\lambda\varphi\phi_y)\beta_\pi^2}{1+\gamma\varphi\phi_\pi+\varphi\phi_y}, \qquad \text{(A.11)}$$

$$\lambda_1\lambda_2=\frac{\lambda\beta_y^2\beta_\pi^2}{1+\gamma\varphi\phi_\pi+\varphi\phi_y}. \qquad \text{(A.12)}$$

From these expressions, it is easy to see that $G_1(\beta_y,\beta_\pi)$ and $G_2(\beta_y,\beta_\pi)$ are analytic functions with respect to β_y and β_π, *independent* of $\boldsymbol{\alpha}$.

Finally, the covariance between output gap and inflation is given as

$$\begin{aligned}
E(y_t\pi_t) &= \big(-\sigma_y^2\gamma\big(-(1+\gamma\varphi\phi_\pi+\varphi\phi_y)\big(1+\gamma\varphi\phi_\pi+\varphi\phi_y+\beta_y^2\rho\big) \\
&\qquad+\beta_\pi^4\lambda\big(1+\gamma\varphi\phi_\pi+\varphi\phi_y+\beta_y^2\rho\big)(\lambda+\gamma\varphi+\lambda\varphi\phi_y)
\end{aligned}$$

$$
\begin{aligned}
&+\beta_\pi^2\rho\big[\beta_y^4\lambda+\beta_y^2\lambda\rho(1+\gamma\varphi\phi_\pi+\varphi\phi_y)+\gamma\varphi(-1+\lambda\phi_\pi)(1+\gamma\varphi\phi_\pi+\varphi\phi_y)\big]\\
&-\beta_y^2\beta_\pi^6\lambda^2\rho\big(\beta_y^2\lambda+\rho(\lambda+\gamma\varphi+\lambda\varphi\phi_y)\big)\big)\\
&+\sigma_\pi^2\varphi\big(-\phi_\pi(1+\gamma\varphi\phi_\pi+\varphi\phi_y)\big[-\beta_y^4-\beta_y^2\gamma\rho\varphi\phi_\pi+(1+\varphi\phi_y)\\
&\cdot(1+\gamma\varphi\phi_\pi+\varphi\phi_y)\big]\\
&+\beta_y^2\beta_\pi^6\lambda\rho\big[-\gamma\varphi\big(-\beta_y^2+\rho+\rho\varphi\phi_y\big)+\lambda\rho\big(\beta_y^4-(1+\varphi\phi_y)^2\big)\big]\\
&+\beta_\pi^4\big(\gamma\varphi\big(1-\beta_y^2\rho+\varphi\phi_y\big)(1+\gamma\varphi\phi_\pi+\varphi\phi_y)+\lambda\big(-1+\beta_y^2\rho-\varphi(\gamma\phi_\pi+\phi_y)\big)\\
&\cdot\big(\beta_y^4-(1+\varphi\phi_y)^2\big)+\beta_y^2\lambda^2\rho\phi_\pi\big(-\beta_y^4+(1+\varphi\phi_y)^2\big)\big)+\beta_\pi^2\rho\big[-\beta_y^6\lambda\rho\phi_\pi\\
&+\beta_y^2\lambda\rho\phi_\pi(1+\varphi\phi_y)(1+\gamma\varphi\phi_\pi+\varphi\phi_y)-(-1+\lambda\phi_\pi)\\
&\cdot(1+\varphi\phi_y)^2(1+\gamma\varphi\phi_\pi+\varphi\phi_y)+\beta_y^4\big(-1-\varphi(\gamma\phi_\pi+\varphi_y)+\lambda(\phi_\pi+\varphi\phi_\pi\phi_y)\big)\big]\big)\big)\\
&/\big(\big(-1+\rho^2\big)\big(-1+\beta_y^2\beta_\pi^2\lambda-\varphi(\gamma\phi_\pi+\phi_y)\big)\big(1+\beta_y^2\rho\big(-1+\beta_\pi^2\lambda\rho\big)+\gamma\varphi\phi_\pi\\
&+\varphi\phi_y-\beta_\pi^2\rho(\lambda+\gamma\varphi+\lambda\varphi\phi_y)\big)\big(\beta_y^4\big(-1+\beta_\pi^4\lambda^2\big)+2\beta_y^2\beta_\pi^2\gamma\varphi(-1+\lambda\phi_\pi)\\
&+(1+\gamma\varphi\phi_\pi+\varphi\phi_y)^2-\beta_\pi^4(\lambda+\gamma\varphi+\lambda\varphi\phi_y)^2\big)\big).
\end{aligned}
\tag{A.13}
$$

A.6 *E-stability of BLE for a baseline NK model*

This Appendix shows the E-stability condition in Corollary 2. Based on Proposition 2, we only need to show that both of the eigenvalues of $(\boldsymbol{I}-\boldsymbol{B}\boldsymbol{\beta}^2)^{-1}(\boldsymbol{B}-\boldsymbol{I})$ have negative real parts if $\gamma(\phi_\pi-1)+(1-\lambda)\phi_y>0$.

The characteristic polynomial of $(\boldsymbol{I}-\boldsymbol{B}\boldsymbol{\beta}^2)^{-1}(\boldsymbol{B}-\boldsymbol{I})$ is given by $h(\nu)=\nu^2-c_1\nu+c_2$, where c_1 is the trace and c_2 is the determinant of matrix $(\boldsymbol{I}-\boldsymbol{B}\boldsymbol{\beta}^2)^{-1}(\boldsymbol{B}-\boldsymbol{I})$. Direct calculation shows that

$$
c_1=\frac{-(1-\lambda)\big(1-\beta_y^2\big)-2\varphi(\gamma\phi_\pi+\phi_y)+\varphi(\gamma+\lambda\phi_y)\big(1+\beta_\pi^2\big)}{\triangle(1+\gamma\varphi\phi_\pi+\varphi\phi_y)},
\tag{A.14}
$$

$$
c_2=\frac{\varphi\big[\gamma(\phi_\pi-1)+(1-\lambda)\phi_y\big]}{\triangle(1+\gamma\varphi\phi_\pi+\varphi\phi_y)},
\tag{A.15}
$$

where $\triangle=\frac{(1-\beta_y^2)(1-\lambda\beta_\pi^2)+\gamma\varphi\phi_\pi+\varphi\phi_y-(\gamma\varphi+\lambda\varphi\phi_y)\beta_\pi^2}{1+\gamma\varphi\phi_\pi+\varphi\phi_y}$.

Both of the eigenvalues of $(\boldsymbol{I}-\boldsymbol{B}\boldsymbol{\beta}^2)^{-1}(\boldsymbol{B}-\boldsymbol{I})$ have negative real parts if and only if $c_1<0$ and $c_2>0$ (these conditions are obtained by applying the *Routh–Hurwitz criterion theorem* (see Brock and Malliaris (1989)). If $\gamma(\phi_\pi-1)+(1-\lambda)\phi_y>0$, from Appendix A.4, it is easy to see that $\triangle>0$. Furthermore,

$$
c_1\le\frac{-2\varphi\big[(\gamma(\phi_\pi-1)+(1-\lambda)\phi_y\big]}{\triangle(1+\gamma\varphi\phi_\pi+\varphi\phi_y)}<0,\qquad c_2>0.
$$

Appendix B: Iterative E-stability algorithm for BLE

This section discusses the Iterative E-stability algorithm used in the approximation of BLE. The first-order autocorrelation coefficients $\boldsymbol{\beta}^*$ in a BLE are functions in terms of the

Algorithm 1: Approximation of a BLE using Iterative E-stability

Denote the set of structural parameters by $\boldsymbol{\mu}$, and the first-order autocorrelation function for a given $\boldsymbol{\mu}$ by $G(\boldsymbol{\beta}^{(k)}, \boldsymbol{\mu})$.

- Step (0): Initialize the vector of learning parameters at $\boldsymbol{\beta}^{(0)}$.
- Step (I): At each iteration k, using the first-order autocorrelation functions, update the vector of learning parameters as

$$\boldsymbol{\beta}^{(k)} = G\big(\boldsymbol{\beta}^{(k-1)}, \boldsymbol{\mu}\big), \tag{B.1}$$

  where $G(\boldsymbol{\beta}^{(k-1)}, \boldsymbol{\mu})$ is known from iteration $k-1$.
- Step (II): Terminate if $\|\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(k-1)}\|_p < \epsilon$, for a small scalar $\epsilon > 0$[4] and a suitable norm distance $\|\cdot\|_p$, otherwise repeat Step (I).

structural parameters $\boldsymbol{\mu}$, which satisfy the nonlinear equilibrium conditions $G(\boldsymbol{\beta}^*, \boldsymbol{\mu}) = \boldsymbol{\beta}^*$ in (2.41). In order to find a BLE for a given $\boldsymbol{\mu}$, we use a simple fixed-point iteration, which is formalized below in Algorithm 1.

A BLE $(\mathbf{0}, \boldsymbol{\beta^*})$ is locally stable under (B.1) if all eigenvalues of $DG_{\boldsymbol{\beta}}(\boldsymbol{\beta^*})$ lie inside the unit circle. Then the equilibrium is said to be *iteratively E-stable*. When Algorithm 1 terminates for some K at a small prespecified ϵ, we say that it has converged to $\boldsymbol{\beta}^{(K)}$. Note that if Algorithm 1 converges, it converges to an approximate BLE since

$$\left\|\boldsymbol{\beta}^{(K+1)} - \boldsymbol{\beta}^{(K)}\right\| < \epsilon \quad \Rightarrow \quad \left\|\boldsymbol{G}\big(\boldsymbol{\beta}^{(K)}\big) - \boldsymbol{\beta}^{(K)}\right\| < \epsilon \quad \Rightarrow \quad \boldsymbol{G}\big(\boldsymbol{\beta}^{(K)}\big) \approx \boldsymbol{\beta}^{(K)}.$$

Given a vector of initial values for the first-order autocorrelation coefficients of the forward-looking variables, we use $N = 200$ iterations under (2.41) for each parameter draw and use the resulting fixed point as an approximate BLE. The algorithm typically takes less than 50 iterations to converge for a given parameter draw, hence $N = 200$ is a conservative value.

Appendix C: Reduced-form matrices

Recall that we consider linear DSGE models in the following general form, as described in Section 2.1:

$$\boldsymbol{x}_t = \boldsymbol{F}\big(\boldsymbol{x}^e_{t+1}, \boldsymbol{u}_t, \boldsymbol{v}_t\big) = \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{x}^e_{t+1} + \boldsymbol{b}_2 \boldsymbol{x}_{t-1} + \boldsymbol{b}_3 \boldsymbol{u}_t + \boldsymbol{b}_4 \boldsymbol{v}_t, \tag{C.1}$$

$$\boldsymbol{u}_t = \boldsymbol{a} + \boldsymbol{\rho} \boldsymbol{u}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{C.2}$$

This Appendix derives the reduced-form matrices $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{B}}$, and $\widehat{\boldsymbol{C}}$ for the equilibrium models, and $\widehat{\boldsymbol{A}_t}$, $\widehat{\boldsymbol{B}_t}$, and $\widehat{\boldsymbol{C}_t}$ for the learning models. The reduced-form matrices are used for the Kalman filter discussed in Appendix D.

[4] In the remainder of this paper, we use the common L^1-Norm as our norm distance, that is, $\|\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(k-1)}\|_p = \sum_{j=1}^{N} |\beta_j^{(k)} - \beta_j^{(k-1)}|$.

C.1 *Equilibrium models*

REE Using the MSV solution, expectations under REE are given by

$$E_t \boldsymbol{x}_{t+1} = \boldsymbol{c}_0 + \boldsymbol{c}_2 \boldsymbol{a} + \boldsymbol{c}_1 \boldsymbol{x}_t + \boldsymbol{c}_2 \boldsymbol{\rho} \boldsymbol{u}_t. \tag{C.3}$$

Plugging back into (C.1) and rewriting yields

$$\boldsymbol{X}_t = \widehat{\boldsymbol{A}} + \widehat{\boldsymbol{B}} \boldsymbol{X}_{t-1} + \widehat{\boldsymbol{C}} \boldsymbol{\eta}_t, \tag{C.4}$$

where $\widehat{\boldsymbol{A}} = \begin{bmatrix} I - \boldsymbol{b}_1 \boldsymbol{c}_1 & -(\boldsymbol{b}_1 \boldsymbol{c}_2 \boldsymbol{\rho} + \boldsymbol{b}_3) \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{c}_0 + \boldsymbol{b}_1 \boldsymbol{c}_2 \boldsymbol{a} \\ \boldsymbol{a} \end{bmatrix}$, $\widehat{\boldsymbol{B}} = \begin{bmatrix} I - \boldsymbol{b}_1 \boldsymbol{c}_1 & -(\boldsymbol{b}_1 \boldsymbol{c}_2 \boldsymbol{\rho} + \boldsymbol{b}_3) \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_2 & 0 \\ 0 & \boldsymbol{\rho} \end{bmatrix}$, $\widehat{\boldsymbol{C}} = \begin{bmatrix} I - \boldsymbol{b}_1 \boldsymbol{c}_1 & -(\boldsymbol{b}_1 \boldsymbol{c}_2 \boldsymbol{\rho} + \boldsymbol{b}_3) \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_4 & 0 \\ 0 & I \end{bmatrix}$, $\boldsymbol{X}_t' = [\boldsymbol{x}_t', \boldsymbol{u}_t']$, $\boldsymbol{\eta}_t' = [\boldsymbol{v}_t', \boldsymbol{\varepsilon}_t']$.

BLE Expectations under BLE are given by

$$E_t \boldsymbol{x}_{t+1} = \boldsymbol{\alpha}^* + \boldsymbol{\beta}^{*2} \big(\boldsymbol{x}_{t-1} - \boldsymbol{\alpha}^* \big). \tag{C.5}$$

Plugging back into (C.1) and rewriting yields

$$\boldsymbol{X}_{t+1} = \widehat{\boldsymbol{A}} + \widehat{\boldsymbol{B}} \boldsymbol{X}_{t-1} + \widehat{\boldsymbol{C}} \boldsymbol{\eta}_t, \tag{C.6}$$

where $\widehat{\boldsymbol{A}} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{\alpha}^* - \boldsymbol{b}_1 \boldsymbol{\beta}^{*2} \boldsymbol{\alpha}^* \\ \boldsymbol{a} \end{bmatrix}$, $\widehat{\boldsymbol{B}} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_1 \boldsymbol{\beta}^{*2} + \boldsymbol{b}_2 & 0 \\ 0 & \boldsymbol{\rho} \end{bmatrix}$, $\widehat{\boldsymbol{C}} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_4 & 0 \\ 0 & I \end{bmatrix}$.

C.2 *Learning models*

SAC-learning Expectations under SAC-learning are given by

$$E_t \boldsymbol{x}_{t+1} = \boldsymbol{\alpha}_{t-1} + \boldsymbol{\beta}_{t-1}^2 (\boldsymbol{x}_{t-1} - \boldsymbol{\alpha}_{t-1}). \tag{C.7}$$

Plugging back into (C.1) and rewriting yields

$$\boldsymbol{X}_t = \widehat{\boldsymbol{A}}_{t-1} + \widehat{\boldsymbol{B}}_{t-1} \boldsymbol{X}_{t-1} + \widehat{\boldsymbol{C}}_{t-1} \boldsymbol{\eta}_t, \tag{C.8}$$

where $\widehat{\boldsymbol{A}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{\alpha}_{t-1} - \boldsymbol{b}_1 \boldsymbol{\beta}_{t-1}^2 \boldsymbol{\alpha}_{t-1} \\ \boldsymbol{a} \end{bmatrix}$, $\widehat{\boldsymbol{B}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_1 \boldsymbol{\beta}_{t-1}^2 + \boldsymbol{b}_2 & 0 \\ 0 & \boldsymbol{\rho} \end{bmatrix}$, $\widehat{\boldsymbol{C}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_4 & 0 \\ 0 & I \end{bmatrix}$.

AR(2)-learning Expectations under AR(2)-learning are given by

$$E_t \boldsymbol{x}_{t+1} = \alpha_{t-1} + \boldsymbol{\beta}_{1,t-1} \boldsymbol{x}_{t-1} + \boldsymbol{\beta}_{2,t-1} \boldsymbol{x}_{t-2}. \tag{C.9}$$

Plugging back into (C.1) and rewriting yields

$$\boldsymbol{X}_t = \tilde{\boldsymbol{A}}_{t-1} + \tilde{\boldsymbol{B}}_{t-1} \boldsymbol{X}_{t-1} + \tilde{\boldsymbol{C}}_{t-1} \boldsymbol{X}_{t-2} + \tilde{\boldsymbol{D}}_{t-1} \boldsymbol{\eta}_t, \tag{C.10}$$

where $\tilde{\boldsymbol{A}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{\alpha}_{t-1} \\ \boldsymbol{a} \end{bmatrix}$, $\tilde{\boldsymbol{B}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_1 \boldsymbol{\beta}_{1,t-1} + \boldsymbol{b}_2 & 0 \\ 0 & \boldsymbol{\rho} \end{bmatrix}$, $\tilde{\boldsymbol{C}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \times \begin{bmatrix} \boldsymbol{b}_1 \boldsymbol{\beta}_{2,t-1} & 0 \\ 0 & 0 \end{bmatrix}$, $\tilde{\boldsymbol{D}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_4 & 0 \\ 0 & I \end{bmatrix}$.

This can be further rewritten as

$$\tilde{\boldsymbol{X}}_t = \widehat{\boldsymbol{A}}_{t-1} + \widehat{\boldsymbol{B}}_{t-1} \tilde{\boldsymbol{X}}_{t-1} + \widehat{\boldsymbol{C}}_{t-1} \boldsymbol{\eta}_t, \tag{C.11}$$

where $\tilde{\boldsymbol{X}}_t = \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \end{bmatrix}$, $\widehat{\boldsymbol{A}}_{t-1} = \begin{bmatrix} I & -\tilde{\boldsymbol{B}}_{t-1} \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\boldsymbol{A}}_{t-1} \\ 0 \end{bmatrix}$, $\widehat{\boldsymbol{B}}_{t-1} = \begin{bmatrix} I & -\tilde{\boldsymbol{B}}_{t-1} \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} 0 & \tilde{\boldsymbol{C}}_{t-1} \\ I & 0 \end{bmatrix}$, $\tilde{\boldsymbol{C}}_{t-1} = \begin{bmatrix} I & -\tilde{\boldsymbol{B}}_{t-1} \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\boldsymbol{D}}_{t-1} \\ 0 \end{bmatrix}$.

Pseudo-MSV and VAR(1)-learning Expectations under pseudo-MSV and VAR(1)-learning are given by

$$E_t \boldsymbol{x}_{t+1} = \gamma_{0,t-1} + \gamma_{1,t-1}\boldsymbol{x}_{t-1} + \gamma_{2,t-1}\boldsymbol{\rho}\boldsymbol{u}_{t-1}. \tag{C.12}$$

Plugging back into (C.1) and rewriting yields

$$\boldsymbol{X}_t = \widehat{\boldsymbol{A}}_{t-1} + \widehat{\boldsymbol{B}}_{t-1}\boldsymbol{X}_{t-1} + \widehat{\boldsymbol{C}}_{t-1}\boldsymbol{\eta}_t, \tag{C.13}$$

where $\widehat{\boldsymbol{A}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_0 + \boldsymbol{b}_1\gamma_{0,t-1} \\ \boldsymbol{a} \end{bmatrix}$, $\widehat{\boldsymbol{B}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_1\gamma_{1,t-1} + \boldsymbol{b}_1 & \boldsymbol{b}_2\gamma_{2,t-1}\boldsymbol{\rho} \\ 0 & \boldsymbol{\rho} \end{bmatrix}$, $\widehat{\boldsymbol{C}}_{t-1} = \begin{bmatrix} I & -\boldsymbol{b}_3 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_4 & 0 \\ 0 & I \end{bmatrix}$.

VAR(1)-learning is the special case with $\gamma_{2,t-1} = 0$.

Appendix D: Kalman filter and estimation

This section describes the Kalman filter used in the estimation of equilibrium and adaptive learning models.

D.1 *Kalman filter for equilibrium models*

For the REE model, expectations $\boldsymbol{x}^e_{t+1}$ are pinned down by the equilibrium conditions (2.15)–(2.18). The fixed point of the system (2.15)–(2.18) is computed using standard methods, as in Uhlig (1995).

For the BLE model, expectations $\boldsymbol{x}^e_{t+1}$ are pinned down by the equilibrium conditions (consistency requirements) in (2.27). The fixed point associated with (2.27) is approximated by the iterative E-stability algorithm in (2.41), which converges to the corresponding equilibrium persistence values $\boldsymbol{\beta}^*$ (see Appendix B).

Both REE and BLE models can be written in the following recursive form:

$$\boldsymbol{X}_t = \widehat{\boldsymbol{A}} + \widehat{\boldsymbol{B}}\boldsymbol{X}_{t-1} + \widehat{\boldsymbol{C}}\boldsymbol{\eta}_t, \tag{D.1}$$

with $\boldsymbol{X}'_t = [\boldsymbol{x}'_t, \boldsymbol{u}'_t]$, $\boldsymbol{\eta}'_t = [\boldsymbol{v}'_t, \boldsymbol{\varepsilon}'_t]$ and $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{B}}$, and $\widehat{\boldsymbol{C}}$ matrices of structural parameters. REE and BLE models are characterized by different matrices $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{C}}$, as described in Appendix C.1. Given the linear structure of both models, the likelihood function can be evaluated using a standard Kalman filter.

We denote the initial state vector and state covariance matrix by $\boldsymbol{X}_{0|0}$ and $\boldsymbol{P}_{0|0}$, respectively, while L refers to the number of shocks. The measurement equations are denoted by $\boldsymbol{Y}_t = \bar{\phi} + \phi_1\boldsymbol{X}_t$, with $\boldsymbol{Y}_t$ denoting the vector of observable variables.[5] For every

[5]The SW07 model considered in this paper consists of GDP, consumption, wage and investment growth, CPI inflation, hours worked, and interest rates.

period t, the Kalman-filter recursions are given as follows:

$$\begin{cases} \boldsymbol{X}_{t|t-1} = \widehat{\boldsymbol{A}} + \widehat{\boldsymbol{B}}\boldsymbol{X}_{t-1|t-1}, \\ \boldsymbol{P}_{t|t-1} = \widehat{\boldsymbol{B}}\boldsymbol{P}_{t-1|t-1}\widehat{\boldsymbol{B}}' + \widehat{\boldsymbol{C}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\widehat{\boldsymbol{C}}', \\ v_t = \boldsymbol{Y}_t - \bar{\phi} - \phi_1 \boldsymbol{X}_{t|t-1}, \\ \Sigma_t = \phi_1 \boldsymbol{P}_{t|t-1}\phi_1', \\ \boldsymbol{X}_{t|t} = \boldsymbol{X}_{t|t-1} + \boldsymbol{P}_{t|t-1}\phi_1'\Sigma_t^{-1}v_t, \\ \boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1}\phi_1'\Sigma_t^{-1}\phi_1\boldsymbol{P}_{t|t-1}, \\ L(y_t) = -\frac{L}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_t| - \frac{1}{2}\big(v_t'\Sigma_t^{-1}v_t\big). \end{cases} \tag{D.2}$$

D.2 *Kalman filter for learning models*

Learning models can be represented as a recursive linear system after plugging in the expectations

$$\boldsymbol{X}_t = \widehat{\boldsymbol{A}}_{t-1} + \widehat{\boldsymbol{B}}_{t-1}\boldsymbol{X}_{t-1} + \widehat{\boldsymbol{C}}_{t-1}\boldsymbol{\eta}_t, \tag{D.3}$$

with time-varying matrices $\widehat{\boldsymbol{B}}_{t-1}$, $\widehat{\boldsymbol{C}}_{t-1}$, and perceived mean vector $\widehat{\boldsymbol{A}}_{t-1}$, where the time variation comes from agents' PLM coefficients. The coefficients are updated every period using the SAC-learning or constant gain recursive least squares algorithms in (3.2)–(3.6). Given the $t-1$ timing structure discussed in Section 3.1, the learning models admit a conditionally linear structure given the belief coefficients. Accordingly, the likelihood function can be evaluated using the standard Kalman-filter recursions conditional on the belief coefficients.

We denote the initial perceived covariance matrix and initial belief coefficients that appear in the learning algorithms (3.2)–(3.6) by R_0 and θ_0. For every period t, the conditionally linear Kalman-filter recursions are given as follows:

$$\begin{cases} \textbf{Kalman-filter step:} \\ \boldsymbol{X}_{t|t-1} = \widehat{\boldsymbol{A}}_{t-1} + \widehat{\boldsymbol{B}}_{t-1}\boldsymbol{X}_{t-1|t-1}, \\ \boldsymbol{P}_{t|t-1} = \widehat{\boldsymbol{B}}_{t-1}\boldsymbol{P}_{t-1|t-1}\widehat{\boldsymbol{B}}_{t-1}' + \widehat{\boldsymbol{C}}_{t-1}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\widehat{\boldsymbol{C}}_{t-1}', \\ v_t = \boldsymbol{Y}_t - \bar{\phi} - \phi_1\boldsymbol{X}_{t|t-1}, \\ \Sigma_t = \phi_1\boldsymbol{P}_{t|t-1}\phi_1', \\ \boldsymbol{X}_{t|t} = \boldsymbol{X}_{t|t-1} + \boldsymbol{P}_{t|t-1}\phi_1'\Sigma_t^{-1}v_t, \\ \boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1}\phi_1'\Sigma_t^{-1}\phi_1\boldsymbol{P}_{t|t-1}, \\ L(y_t|R_{t-1}, \theta_{t-1}) \\ \quad = -\frac{L}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_t| - \frac{1}{2}\big(v_t'\Sigma_t^{-1}v_t\big), \end{cases} \quad \begin{cases} \textbf{Learning step:} \\ R_t = T_1(R_{t-1}, \boldsymbol{X}_{t|t}), \\ \theta_t = T_2(\theta_{t-1}, \boldsymbol{X}_{t|t}, R_t), \end{cases} \tag{D.4}$$

where T_1 and T_2 correspond to the updating equations for belief coefficients, as outlined in (3.2)–(3.6). In this context, we treat the learning models as a *temporary equilibrium* system, where the matrices $\widehat{\boldsymbol{B}}_{t-1}$ and $\widehat{\boldsymbol{C}}_{t-1}$ and vector $\widehat{\boldsymbol{A}}_{t-1}$ are updated every period. The Kalman-filter step treats the model as a linear system for a given θ_t and returns the likelihood function and state variables $\boldsymbol{X}_{t|t}$. The state variables are used as an input to update the belief coefficients in the learning step.

D.3 *Estimation*

The model parameters are estimated using standard Bayesian likelihood methods. This consists of a posterior mode search step and a Monte Carlo Markov Chain (MCMC) step, which are summarized below.

1. Prior distributions of the estimated parameters are specified. For the SW07 model used in this paper, the prior distributions are summarized in Table E.1. We denote the prior distribution function by $p(\boldsymbol{\mu})$.

2. The likelihood function $p(\boldsymbol{\mu}) * L(y_t|\boldsymbol{\mu})$ is maximized using standard iterative gradient descent algorithms. We use the *csminwel* algorithm (Sims (1999)) available in MATLAB software.

    - The algorithm iteratively updates the parameters $\boldsymbol{\mu}$ until convergence. We denote each iteration of parameter draws by $\boldsymbol{\mu}^n$.

    - Equilibrium models: For every parameter draw $\boldsymbol{\mu}^n$, the equilibrium is calculated by finding the fixed point of (2.15)–(2.18) and (2.27) for REE and BLE models, respectively. For the REE model the fixed point is found by using Uhlig's method (1995). For the BLE model, the iterative E-stability algorithm in (B.1) with 200 iterations is used. Given the reduced-form matrices $\widehat{\boldsymbol{A}}(\boldsymbol{\mu}^n)$, $\widehat{\boldsymbol{B}}(\boldsymbol{\mu}^n)$ and $\widehat{\boldsymbol{C}}(\boldsymbol{\mu}^n)$, the likelihood function $p(\boldsymbol{\mu}^n) * L(y_t|\boldsymbol{\mu}^n)$ is computed using the Kalman- filter recursions in (D.2).

    - Learning models: For every parameter draw $\boldsymbol{\mu}^n$, the reduced-form matrices $\widehat{\boldsymbol{A}}_{t-1}(\boldsymbol{\mu}^n)$, $\widehat{\boldsymbol{B}}_{t-1}(\boldsymbol{\mu}^n)$ and $\widehat{\boldsymbol{C}}_{t-1}(\boldsymbol{\mu}^n)$ are recalculated at every period t of the Kalman filter in (D.4). Given the reduced-form matrices, the likelihood function $p(\boldsymbol{\mu}^n) * L(y_t|\boldsymbol{\mu}^n)$ is computed.

    - The parameter values $\boldsymbol{\mu}^n$ are iteratively updated until the likelihood function $p(\boldsymbol{\mu}^n) * L(y_t|\boldsymbol{\mu}^n)$ converges. The optimized parameter values $\boldsymbol{\mu}^*$ are referred to as the posterior mode.

3. For each model, the MCMC algorithm is initialized at $(\boldsymbol{\mu}^*, c\boldsymbol{\Sigma}^*_{\boldsymbol{\mu}})$, where $\boldsymbol{\Sigma}^*_{\boldsymbol{\mu}}$ denotes the covariance matrix of $\boldsymbol{\mu}^*$. c is a scaling coefficient tuned to obtain an average acceptance ratio between 30 and 45%. For each model, we use two parallel chains with 500,000 draws and discard the first half as a burn-in sample. The second half of the chains is used to compute the posterior moments reported in Tables 1 and 3.

   For the BLE model, the initial values of the first-order autocorrelation coefficients of the AR(1) beliefs $\boldsymbol{\beta}^{(0)}$, that is, Step (0) of Algorithm 1 in Appendix B, are fixed prior to the estimation. For forward-looking variables that are observable, that is, inflation π_t and hours worked l_t, the initial values are set to the corresponding first-order sample-autocorrelation over the estimation period. For the remaining latent forward-looking variables, we take the unconditional first-order autocorrelations implied by the estimated REE. We keep the initial values $\boldsymbol{\beta}^{(\mathbf{0})}$ fixed at these values for all parameter draws $\boldsymbol{\mu}^n$. Parameter draws where the fixed-point iteration fails to converge to a stationary equilibrium are discarded.

Appendix E: The Smets–Wouters 2007 model

E.1 *Model descriptions*

The model consists of 13 equations linearized around the steady-state growth path, supplemented with seven exogenous structural shocks. We deviate from the benchmark model by slightly restricting the parameter space of the model, where we assume all shocks follow an AR(1) process.[6] In this section, we briefly outline the resulting linearized model economy that is used in our estimation. To start with the demand side of the economy, the aggregate resource constraint is given by

$$\begin{cases} \tilde{y}_t = c_y c_t + i_y i_t + z_y z_t + \epsilon_t^g, \\ \epsilon_t^g = \rho_g \epsilon_{t-1}^g + \eta_t^g, \end{cases} \tag{E.1}$$

where $\tilde{y}_t$, c_t, i_t, and z_t are the output, consumption, investment, and capital utilization rate, respectively, while c_y, i_y, and z_y are the steady-state shares in output of the respective variables. The second equation in (E.1) defines the exogenous spending shock ϵ_t^g, where η_t^g is an i.i.d-normal disturbance for spending. The consumption Euler equation is given by

$$\begin{cases} c_t = c_1 c_{t-1} + (1-c_1)\mathbb{E}_t c_{t+1} + c_2(l_t - \mathbb{E}_t l_{t+1}) - c_3(r_t - \mathbb{E}_t \pi_{t+1}) + \epsilon_t^b, \\ \epsilon_t^b = \rho_b \epsilon_{t-1}^b + \eta_t^b, \end{cases} \tag{E.2}$$

with $c_1 = \frac{\lambda}{\gamma}/(1+\frac{\lambda}{\gamma})$, $c_2 = (\sigma_c - 1)(w_{ss}l_{ss}/c_{ss})/(\sigma_c(1+\frac{\lambda}{\gamma}))$, $c_3 = (1-\frac{\lambda}{\gamma})/((1+\frac{\lambda}{\gamma})\sigma_c)$, where λ, γ, and σ_c denote the habit formation in consumption, steady-state growth rate, and the elasticity of intertemporal substitution, respectively, while x_{ss} corresponds to the steady-state level of a given variable x. The equation implies that current consumption is a weighted average of the past and expected future consumption, expected growth in hours worked and the ex ante real interest rate. ϵ_t^b corresponds to the risk premium shock modeled as an AR(1) process, where η_t^b is an i.i.d-normal disturbance. Next, the investment Euler equation is defined as

$$\begin{cases} i_t = i_1 i_{t-1} + (1-i_1)\mathbb{E}_t i_{t+1} + i_2 q_t + \epsilon_t^i, \\ \epsilon_t^i = \rho_i \epsilon_{t-1}^i + \eta_t^i, \end{cases} \tag{E.3}$$

with $i_1 = \frac{1}{1+\bar{\beta}\gamma}$, $i_2 = \frac{1}{(1+\bar{\beta}\gamma)(\gamma^2\phi)}$, where $\bar{\beta} = \beta\gamma^{-\sigma_c}$, ϕ is the steady-state elasticity of capital adjustment cost and β is the HH discount factor. q_t denotes the real value of existing

[6]In particular, the benchmark model has more structure on the exogenous shocks, where the mark-up shocks each follow an ARMA(1,1) process and the technology and government spending shocks follow a VAR(1) process. We refer the reader to SW for more details about the microfoundations of the benchmark model.

capital stock. Similar to the consumption Euler, the equation implies that investment is a weighted average of past and expected future consumption, as well as the real value of existing capital stock. ϵ_t^i represents the AR(1) investment shock, where η_t^i is an i.i.d-normal disturbance. The value of the capital-arbitrage equation is given by

$$q_t = q_1 \mathbb{E}_t q_{t+1} + (1-q_1)\mathbb{E}_t r_{t+1}^k - (r_t - \mathbb{E}_t \pi_{t+1}) + \frac{1}{c_3}\epsilon_t^b, \tag{E.4}$$

with $q_1 = \bar{\beta}(1-\delta)$, implying the real value of capital stock is a weighted average of its expected future value and expected real rental rate on capital, net of ex ante real interest rate and the risk premium shock. The production function is characterized as

$$\begin{cases} \tilde{y}_t = \phi_p\big(\alpha k_t^s + (1-\alpha) l_t + \epsilon_t^a\big), \\ \epsilon_t^a = \rho_a \epsilon_{t-1}^a + \eta_t^a, \end{cases} \tag{E.5}$$

where k_t^s denotes the capital services used in production, α is the share of capital in production, and ϕ_p is (one plus) the share of fixed costs in production. ϵ_t^a denotes the AR(1) total factor productivity shock. Capital is assumed to be the sum of the previous amount of capital services used and the degree of capital utilization. Hence,

$$k_t^s = k_{t-1} + z_t. \tag{E.6}$$

Moreover, the degree of capital utilization is a positive function of the degree of rental rate, $z_t = z_1 r_t^k$, with $z_1 = \frac{1-\psi}{\psi}$, ψ being the elasticity of the capital utilization adjustment cost. Next, the equation for installed capital is given by

$$k_t = k_1 k_{t-1} + (1-k_1) i_t + k_2 \epsilon_t^i, \tag{E.7}$$

with $k_1 = \frac{1-\delta}{\gamma}$, $k_2 = (1 - \frac{1-\delta}{\gamma})(1+\bar{\beta}\gamma)\gamma^2\phi$. The price mark-up equation is given by

$$\mu_t^p = \alpha\big(k_t^s - l_t\big) + \epsilon_t^a - w_t, \tag{E.8}$$

which means the price mark-up μ_t^p is the marginal product of the labor net of the current wage. The NKPC is characterized as

$$\begin{cases} \pi_t = \pi_1 \mathbb{E}_t \pi_{t+1} - \pi_2 \mu_t^p + \epsilon_t^p, \\ \epsilon_t^p = \rho_p \epsilon_{t-1}^p + \eta_t^p, \end{cases} \tag{E.9}$$

with $\pi_1 = \bar{\beta}\gamma$, $\pi_2 = (1-\beta\gamma\xi_p)(1-\xi_p)/[\xi_p((\phi_p - 1)\epsilon_p + 1)]$, where ξ_p corresponds to the degree of price stickiness, while ϵ_p denotes the Kimball goods market aggregator. The equation implies that current inflation is determined by the expected future inflation, the price mark-up, and the AR(1) price mark-up shock ϵ_t^p, where η_t^p is an i.i.d-normal disturbance. The rental rate of capital is given by

$$r_t^k = -(k_t - l_t) + w_t, \tag{E.10}$$

which implies the rental rate of capital is decreasing in the capital-labor ratio and increasing in the real wage. The wage mark-up is given as the real wages net of the marginal rate of substitution between working and consuming. Hence,

$$\mu_t^w = w_t - (\sigma_l l_t + \frac{1}{1-\lambda/\gamma}\left(c_t - \frac{\lambda}{\gamma}c_{t-1}\right), \tag{E.11}$$

where σ_l denotes the elasticity of labor supply. The real wage equation is given by

$$\begin{cases} w_t = w_1 w_{t-1} + (1-w_1)(\mathbb{E}_t w_{t+1} + \mathbb{E}_t \pi_{t+1}) - w_2 \mu_t^w + \epsilon_t^w, \\ \epsilon_t^w = \rho_w \epsilon_{t-1}^w + \eta_t^w, \end{cases} \tag{E.12}$$

with $w_1 = 1/(1+\bar{\beta}\gamma)$ and $w_2 = ((1-\bar{\beta}\gamma\xi_w)(1-\xi_w)/(\xi_w(\phi_w - 1)\epsilon_w + 1))$. Hence, the real wage is a weighted average of the past and expected wage, expected inflation, the wage mark-up and the wage mark-up shock ϵ_t^w, where η_t^w is an i.i.d-normal disturbance. Finally, monetary policy is assumed to follow a standard generalized Taylor rule:

$$\begin{cases} r_t = \rho r_{t-1} + (1-\rho)(\phi_\pi \pi_t + \phi_y y_t) + \phi_{\Delta y}(\Delta y_t) + \epsilon_t^r, \\ \epsilon_t^r = \rho_r \epsilon_{t-1}^r + \eta_t^r, \end{cases} \tag{E.13}$$

where y_t denotes the output gap and ϵ_t^r is the AR(1) monetary policy shock, with η_t^r the i.i.d-normal disturbance. Hence, the monetary policy responds with output gap growth on top of inflation and the output gap. In this paper, following the approach in Slobodyan and Wouters (2012), the output gap is defined as the deviation of output from the underlying productivity process, that is, $y_t = \tilde{y}_t - \Phi_p \epsilon_t^a$. The prior distributions used for all estimated parameters are provided in Table E.1.

We use U.S. historical quarterly macroeconomic data for the period 1966:I–2007:IV. The observable variables used in the estimation are the (log-) difference of real GDP (y_t^{data}), real consumption (c_t^{data}), real investment (inv_t^{data}), real wage (w_t^{data}), log hours worked (l_t^{data}), inflation (π_t^{data}), and the federal funds rate (r_t^{data}) for the U.S. economy. The measurement equations are given as

$$\begin{cases} d(\log(y_t^{\text{data}})) = \bar{\gamma} + (y_t - y_{t-1}), \\ d(\log(c_t^{\text{data}})) = \bar{\gamma} + (c_t - c_{t-1}), \\ d(\log(inv_t^{\text{data}})) = \bar{\gamma} + (inv_t - inv_{t-1}), \\ d(\log(w_t^{\text{data}})) = \bar{\gamma} + (w_t - w_{t-1}), \\ \log(l_t^{\text{data}}) = \bar{l} + l_t, \\ (\log(\pi_t^{\text{data}})) = \bar{\pi} + \pi_t, \\ (\log(r_t^{\text{data}})) = \bar{r} + r_t. \end{cases} \tag{E.14}$$

The construction of the time series follow the same steps as in Smets and Wouters (2007).

E.2 *Deviations from the original model*

Our model follows the original Smets and Wouters (2007) structure with minor deviations. First, we use CPI inflation as our inflation measure instead of the GDP deflator

TABLE E.1. Fixed parameters and the prior distributions of the estimated parameters for the Smets–Wouters (2007) model.

Fixed Parameters			
δ	0.025		
ϕ_w	1.5		
g	0.18		
ϵ_p	10		
ϵ_w	10		
Prior	**Distribution**	**Mean**	**Var.**
Parameters related to nominal and real frictions			
ϕ	Normal	4	1.5
σ_c	Normal	1.5	0.375
λ	Beta	0.7	0.1
ξ_w	Beta	0.5	0.1
σ_l	Normal	2	0.75
ξ_p	Beta	0.5 (0.75)	0.1 (0.05)
ψ	Beta	0.5	0.15
ϕ_p	Normal	1.25	0.125
ι_p	Normal	0.5	0.15
ι_w	Normal	0.5	0.15
Policy related parameters			
ϕ_π	Normal	1.5	0.25
ρ	Beta	0.75	0.1
ϕ_y	Normal	0.125	0.05
$\phi_{\Delta y}$	Normal	0.125	0.05
Steady-state related parameters			
$\bar{\pi}$	Gamma	0.625	0.1
$\bar{\beta}$	Gamma	0.25	0.1
$\bar{l}$	Normal	0	2
$\bar{\gamma}$	Normal	0.4	0.1
α	Normal	0.3	0.05
Parameters related to shock persistence			
ρ_a	Beta	0.5	0.2
ρ_b	Beta	0.5	0.2
ρ_g	Beta	0.5	0.2
ρ_i	Beta	0.5	0.2
ρ_r	Beta	0.5	0.2
ρ_p	Beta	0.5	0.2
ρ_w	Beta	0.5	0.2
ρ_{ga}	Beta	0.5	0.2
Shock variance parameters			
η_a	Inv. Gamma	0.1	2
η_b	Inv. Gamma	0.1	2
η_g	Inv. Gamma	0.1	2
η_i	Inv. Gamma	0.1	2
η_r	Inv. Gamma	0.1	2
η_p	Inv. Gamma	0.1	2
η_w	Inv. Gamma	0.1	2
γ	Gamma	0.035	0.015

used in the original model. Second, we define the output gap in the model as the deviation of output from its natural level based on the productivity process.[7] Third, the observable variables used in the estimation are the (log-) difference of real GDP, real consumption, real investment, real wages, (log-) hours worked, inflation, and the federal funds rate for the U.S. economy. The model structure is the same as the original SW except for three minor deviations. The first is the definition of the output gap: in the original model, this is the deviation of output from its potential level, defined as output in the presence of flexible prices and wages. Instead, we follow Slobodyan and Wouters (2012) and define output gap as the deviation of output from its natural level based on the productivity process.[8] The second deviation involves the exogenous price and wage mark-up shocks, which follow ARMA(1,1) processes in the original model. However, as shown in Slobodyan and Wouters (2012), mark-up shocks are typically reduced to near white noise processes once learning dynamics are introduced. In these cases, the AR(1) and MA(1) parameters are typically locally unidentified. Therefore, we shut off the MA component of these shocks. The third difference pertains to the prior distribution of price stickiness ξ_p, which we tighten from $\xi_p \sim \text{Beta}(0.5, 0.2)$ to $\xi_p \sim \text{Beta}(0.75, 0.05)$. This follows from our observations that the approximation algorithm may fail to find an equilibrium, and thus break down for small values of ξ_p. Further, for the learning models, we have an additional estimated parameter γ, that is, the constant gain value. This is assigned a prior of $\gamma \sim \text{Gamma}(0.035, 0.015)$, which closely follows the assumption in Slobodyan and Wouters (2012). The remainder of the model remains unchanged and consists of 13 equations with 7 forward-looking variables, 7 exogenous shocks, and 7 state variables that enter into the model equations with a lag. There are 35 estimated parameters including the constant gain for the adaptive learning models.

References

Brock, William A. and Anastasios G. Malliaris (1989), *Differential Equations, Stability and Chaos in Dynamic Economics*. Advanced Textbooks in Economics (90A16 BROd). [6]

Elaydi, Saber (2005), *An Introduction to Difference Equations*, third edition. Springer, New York. [4]

Evans, George W. and Seppo Honkapohja (2001), *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton, NJ. [2, 3]

Hommes, Cars and Mei Zhu (2014), "Behavioral learning equilibria." *Journal of Economic Theory*, 150, 778–814. [2]

Lancaster, Peter and Miron Tismenetsky (1985), *The Theory of Matrices: With Applications*. Elsevier. [2]

Magnus, Jan R. and Heinz Neudecker (2019), *Matrix Differential Calculus With Applications in Statistics and Econometrics*. John Wiley & Sons. [2]

[7] This is the approach used in Slobodyan and Wouters (2012).

[8] See Appendix E for further details.

Sims, Christopher (1999), "Matlab optimization software." QM&RBC Codes. [11]

Slobodyan, Sergey and Raf Wouters (2012), "Learning in an estimated medium-scale DSGE model." *Journal of Economic Dynamics and Control*, 36 (1), 26–46. [14, 16]

Smets, Frank and Rafael Wouters (2007), "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97 (3), 586–606. [14, 15]

Uhlig, Harald (1995), "A toolkit for analyzing nonlinear dynamic stochastic models easily." Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics, 55480-0291. [9, 11]

Co-editor Tao Zha handled this manuscript.

Manuscript received 24 January, 2020; final version accepted 26 April, 2023; available online 16 May, 2023.