Supplement to "Monetary policy and long-term interest rates: Online technical appendix"

(*Quantitative Economics*, Vol. 14, No. 2, May 2023, 689–716)

Gianni Amisano
Board of Governors of the Federal Reserve System

Oreste Tristani
European Central Bank and CEPR

This appendix describes some important details regarding aspects of the specification and the estimation of the model used in Amisano and Tristani (2022). The Appendix is available at the following URL: https://sites.google.com/site/gianniamisanowebsite/.

Appendix A: Model details

A.1 *The household problem*

The optimization problem is

$$\max V_t = \left\{(1-\beta)u_t^{1-\psi} + \beta\left(\mathrm{E}_t V_{t+1}^{1-\gamma}\right)^{\frac{1-\psi}{1-\gamma}}\right\}^{\frac{1}{1-\psi}}, \quad \psi, \gamma \neq 1,$$

where u_t is shorthand for $u\{C_t(j) - h\Xi_t C_{t-1}, 1 - N_t(j)\}$, subject to

$$P_t C_t(j) + \mathrm{E}_t Q_{t,t+1} W_{t+1}(j) \leq W_t(j) + w_t(j) N_t(j) + \int_0^1 \Psi_t(i)\,\mathrm{d}i - T_t,$$

and

$$N_t(j) = L_t \left(\frac{w_t(j)}{w_t}\right)^{-\theta_{w,t}},$$

where the choice variables are w_s and c_s.

The Bellman equation is

$$\begin{aligned} J(W_t) = \max&\left\{(1-\beta)u_t^{1-\psi} + \beta\left[\mathrm{E}_t J^{1-\gamma}(W_{t+1})\right]^{\frac{1-\psi}{1-\gamma}}\right\}^{\frac{1}{1-\psi}} \\ &- \Lambda_t\left[P_t C_t + \mathrm{E}_t Q_{t,t+1} W_{t+1} - W_t - w_t N_t - \int_0^1 \Psi_t(i)\,\mathrm{d}i + T_t\right]. \end{aligned}$$

Using the aggregator function,

$$U = \left\{(1-\beta)u_t^{1-\psi} + \beta v_t^{1-\psi}\right\}^{\frac{1}{1-\psi}}$$

Gianni Amisano: gianni.amisano@frb.gov
Oreste Tristani: oreste.tristani@ecb.europa.eu

 https://doi.org/10.3982/QE1287

for $v_t \equiv [\mathrm{E}_t J^{1-\gamma}(W_{t+1}, C_t)]^{\frac{1}{1-\gamma}}$, define

$$U_{u,t} = (1-\beta)\left\{(1-\beta)u_t^{1-\psi} + \beta v_t^{1-\psi}\right\}^{\frac{\psi}{1-\psi}} u_t^{-\psi},$$
$$U_{v,t} = \beta\left\{(1-\beta)u_t^{1-\psi} + \beta v_t^{1-\psi}\right\}^{\frac{\psi}{1-\psi}} v_t^{-\psi}.$$

The FOCs include

$$U_{u,t}u_{c,t} = \Lambda_t P_t,$$
$$u_{N,t}U_{u,t}\frac{\partial N_t(j)}{\partial w_t(j)} = -\Lambda_t\left[N_t(j) + w_t(j)\frac{\partial N_t(j)}{\partial w_t(j)}\right],$$

and state-by-state

$$U_{v,t}\left[\mathrm{E}_t J^{1-\gamma}(W_{t+1})\right]^{\frac{\gamma}{1-\gamma}} J^{-\gamma}(W_{t+1})J_W(W_{t+1}) = \Lambda_t Q_{t,t+1}$$

plus envelope

$$J_W(W_t) = \Lambda_t.$$

Use the shorthand $J_t = J(V_t)$ and $J_{t+1} = J(V_{t+1})$, the FOCs can be rewritten as

$$\frac{\Lambda_t P_t}{u_{c,t}} = U_{u,t},$$
$$\frac{u_{N,t}}{u_{c,t}} = \frac{1-\theta_{w,t}}{\theta_{w,t}}\frac{w_t(j)}{P_t},$$
$$Q_{t,t+1} = U_{v,t}\left[\mathrm{E}_t J_{t+1}^{1-\gamma}\right]^{\frac{\gamma}{1-\gamma}} J_{t+1}^{-\gamma}\frac{\Lambda_{t+1}}{\Lambda_t},$$

or

$$Q_{t,t+1} = \beta\left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}}{J_{t+1}}\right)^{\gamma-\psi}\frac{u_{t+1}^{-\psi}}{u_t^{-\psi}}\frac{u_{c,t+1}}{u_{c,t}}\frac{1}{\Pi_{t+1}}.$$

Using the definition of $\mu_{w,t}$, we obtain, as in the text,

$$-\frac{u_{N,t}}{u_{c,t}} = \mu_{w,t}\frac{w_t(j)}{P_t}$$

and

$$Q_{t,t+1} = \beta\left[\mathrm{E}_t\left(\frac{J_{t+1}}{J_t}\right)^{1-\gamma}\right]^{\frac{\gamma-\psi}{1-\gamma}}\left(\frac{J_{t+1}}{J_t}\right)^{-(\gamma-\psi)}\left(\frac{u_{t+1}}{u_t}\right)^{-\psi}\frac{u_{c,t+1}}{u_{c,t}}\frac{1}{\Pi_{t+1}}.$$

A.2 *Detrending*

Given the stochastic trend B_t, define a detrended variable as $\widetilde{x}_t \equiv x_t/B_t$. It follows that we can rewrite the conditions above as

$$-\frac{\widetilde{u}_{N,t}}{u_{c,t}} = \frac{\theta_{w,t}-1}{\theta_{w,t}}\frac{\widetilde{w}_t(j)}{P_t},$$

$$\begin{aligned}
\widetilde{J}_t^{1-\psi} &= (1-\beta)\widetilde{u}_t^{1-\psi} + \beta\big[\mathrm{E}_t \Xi_{t+1}^{1-\gamma} \widetilde{J}_{t+1}^{1-\gamma}\big]^{\frac{1-\psi}{1-\gamma}}, \\
\widetilde{u}_t &= u\big(\widetilde{C}_t(j) - h\widetilde{C}_{t-1}, 1 - N_t(j)\big), \\
Q_{t,t+1} &= \beta\left(\frac{\big[\mathrm{E}_t \widetilde{J}_{t+1}^{1-\gamma}\Xi_{t+1}^{1-\gamma}\big]^{\frac{1}{1-\gamma}}}{\widetilde{J}_{t+1}\Xi_{t+1}}\right)^{\gamma-\psi}\left(\frac{\widetilde{u}_{t+1}}{\widetilde{u}_t}\right)^{-\psi}\frac{u_{c,t+1}}{u_{c,t}}\frac{1}{\Pi_{t+1}\Xi_{t+1}^{\psi}}.
\end{aligned}$$

A.3 *Expected excess holding period returns*

Define zero-coupon bond prices as

$$\begin{aligned}
B_{t,1} &= \mathrm{E}_t[Q_{t,t+1}], \\
B_{t,2} &= \mathrm{E}_t[Q_{t,t+1}B_{t+1,1}], \\
&\dots \\
B_{t,n} &= \mathrm{E}_t[Q_{t,t+1}B_{t+1,n-1}]
\end{aligned}$$

and note that the 1-period zero-coupon yield on the n-period bond, $I_{n,t}$, is defined as

$$\frac{1}{(I_{t,n})^n} = B_{t,n}$$

or to second order

$$\widehat{i}_{t,n} = -\frac{1}{n}\widehat{b}_{t,n}.$$

Expected holding period returns on a n-period bonds are

$$HPR_{n,t} = \frac{\mathrm{E}_t B_{n-1,t+1}}{B_{n,t}}.$$

Excess holding period returns are defined as $XHPR_{n,t} = HPR_{n,t}/I_t$, where $HPR_{n,t}$ is the return on holding a bond of maturity n for one period given by $HPR_{n,t} = \mathrm{E}_t B_{n-1,t+1}/B_{n,t}$.

A.3.1 *Second-order approximation*

In the text, we use the second-order approximation of expected holding period returns.

From the definition of $HPR_{n,t}$, denoting the approximated holding period return as $\widehat{h}_{n,t}$, we obtain

$$\widehat{h}_{n,t} + \frac{1}{2}\widehat{h}_{n,t}^2 = \mathrm{E}_t\widehat{b}_{n-1,t+1} - \widehat{b}_{n,t} + \frac{1}{2}\mathrm{E}_t\widehat{b}_{n-1,t+1}^2 + \frac{1}{2}\widehat{b}_{n,t}^2 - \widehat{b}_{n,t}\mathrm{E}_t\widehat{b}_{n-1,t+1}.$$

Using first-order terms to evaluate $\widehat{h}_{n,t}^2$, we obtain

$$\widehat{h}_{n,t} = -\widehat{b}_{n,t} + \mathrm{E}_t\widehat{b}_{n-1,t+1} + \frac{1}{2}\mathrm{Var}_t\widehat{b}_{n-1,t+1}.$$

Similarly, for bond prices $B_{t,n} = \mathrm{E}_t[Q_{t,t+1}B_{t+1,n-1}]$ we obtain

$$\widehat{b}_{t,n} = \widehat{b}_{t,1} + \mathrm{E}_t\widehat{b}_{t+1,n-1} + \frac{1}{2}\mathrm{Var}_t\widehat{b}_{t+1,n-1} + \mathrm{Cov}_t[\widehat{b}_{t+1,n-1}, \widehat{q}_{t,t+1}].$$

Using this expression, expected holding period returns can be written as $\widehat{h}_{n,t} = \widehat{i}_t - \mathrm{Cov}_t[\widehat{b}_{t+1,n-1}, \widehat{q}_{t,t+1}]$ and holding period returns in excess of the short rate are

$$\widehat{h}_{n,t} - \widehat{i}_t = -\mathrm{Cov}_t[\widehat{b}_{t+1,n-1}, \widehat{q}_{t,t+1}].$$

Note that we only need a first-order approximation to evaluate the covariance.

The stochastic discount factor $Q_{t,t+1}$ can be rewritten as

$$\begin{aligned}
\widetilde{\Lambda}_t &\equiv \widetilde{u}_t^{-\psi} u_{c,t}, \\
D_t &\equiv \mathrm{E}_t \widetilde{J}_{t+1}^{1-\gamma} \Xi_{t+1}^{1-\gamma}, \\
Q_{t,t+1} &= \beta \frac{D_t^{\frac{\gamma-\psi}{1-\gamma}}}{\widetilde{J}_{t+1}^{\gamma-\psi}} \frac{\widetilde{\Lambda}_{t+1}}{\widetilde{\Lambda}_t} \frac{1}{\Pi_{t+1} \Xi_{t+1}^{\gamma}},
\end{aligned}$$

and approximated as

$$\widehat{q}_{t,t+1} = \Delta \widehat{\widetilde{\lambda}}_{t+1} - \widehat{\pi}_{t+1} - \psi \widehat{\xi}_{t+1} + \frac{\gamma - \psi}{1-\gamma} \widehat{d}_t - (\gamma - \psi) \widehat{\widetilde{j}}_{t+1}.$$

Expanding $\widehat{d}_t$ it follows that, to first order,

$$\widehat{q}_{t,t+1} = \Delta \widehat{\widetilde{\lambda}}_{t+1} - \psi \widehat{\xi}_{t+1} - \widehat{\pi}_{t+1} - (\gamma - \psi)\big(\widehat{\xi}_{t+1} + \widehat{\widetilde{j}}_{t+1} - \mathrm{E}_t[\widehat{\xi}_{t+1} + \widehat{\widetilde{j}}_{t+1}]\big).$$

We therefore obtain

$$\widehat{h}_{n,t} - \widehat{i}_t = -\mathrm{Cov}_t\big[\widehat{b}_{t+1,n-1}, \Delta \widehat{\widetilde{\lambda}}_{t+1} - \psi \widehat{\xi}_{t+1} - \widehat{\pi}_{t+1} - (\gamma - \psi)(\widehat{\xi}_{t+1} + \widehat{\widetilde{j}}_{t+1})\big].$$

We now expand $\widehat{\widetilde{\lambda}}_{t+1}$ for the specific case of the Trabandt and Uhlig (2011) form for temporary utility, which we use in the paper. We have

$$\widetilde{\Lambda}_t = (\widetilde{C}_t - h\widetilde{C}_{t-1})^{-\psi}\big(1 - \eta(1-\psi) N_t^{1+\frac{1}{\phi}}\big)^{\psi},$$

and to first order

$$\widehat{\widetilde{\lambda}}_t = -\frac{\psi}{1-h} \widehat{\widetilde{c}}_t + \psi \frac{h}{1-h} \widehat{\widetilde{c}}_{t-1} - \psi \left(1 + \frac{1}{\phi}\right) \frac{\overline{n}}{1-\overline{n}} \widehat{l}_t$$

for $\overline{n} \equiv \eta(1-\psi) N^{1+\frac{1}{\phi}}$.

Using this expression in the excess holding period return, we obtain

$$\begin{aligned}
&\widehat{h}_{n,t} - \widehat{i}_t \\
&\quad = \mathrm{Cov}_t\Bigg[\widehat{b}_{t+1,n-1}, -\frac{\psi}{1-h} \widehat{\widetilde{c}}_{t+1} - \psi \left(1 + \frac{1}{\phi}\right) \frac{\overline{n}}{1-\overline{n}} \widehat{l}_{t+1} - \psi \widehat{\xi}_{t+1} - \widehat{\pi}_{t+1} \\
&\qquad - (\gamma - \psi)(\widehat{\xi}_{t+1} + \widehat{\widetilde{j}}_{t+1})\Bigg].
\end{aligned}$$

Define the first-order approximation of variable v as $F_v \widehat{\mathbf{x}}_t$. Then (note that we use F_j to denote the first-order approximation of the infinite sum $\widehat{\xi}_{t+1} + \widehat{\widehat{j}}_{t+1}$)

$$\widehat{h}_{n,t} - \widehat{i}_t = \mathrm{Cov}_t\Bigg[F_{B_{n-1}} \widehat{x}_{t+1}, \\ \left(\psi \frac{1}{1-h} F_c + \psi\left(1+\frac{1}{\phi}\right)\frac{\overline{n}}{1-\overline{n}} F_l + \psi F_\xi + F_\pi + (\gamma-\psi) F_j\right)\widehat{x}_{t+1}\Bigg].$$

It follows that

$$\begin{aligned}
&\widehat{h}_{n,t} - \widehat{i}_t \\
&= \mathrm{E}_t\left[F'_{B_{n-1}} \widehat{x}_{t+1}\widehat{x}'_{t+1}\left(\psi \frac{1}{1-h} F_c + \psi\left(1+\frac{1}{\phi}\right)\frac{\overline{n}}{1-\overline{n}} F_l + \psi F_\xi + F_\pi + (\gamma-\psi) F_j\right)'\right] \\
&\quad - \mathrm{E}_t F'_{B_{n-1}} \widehat{x}'_{t+1} \mathrm{E}_t\left[\widehat{\mathbf{x}}'_{t+1}\left(\psi \frac{1}{1-h} F_c + \psi\left(1+\frac{1}{\phi}\right)\frac{\overline{n}}{1-\overline{n}} F_l + \psi F_\xi + F_\pi + (\gamma-\psi) F_j\right)'\right],
\end{aligned}$$

and using the law of motion for $\widehat{x}_{t+1}$,

$$\begin{aligned}
\widehat{h}_{n,t} - \widehat{i}_t &= \widetilde{\sigma}^2 F_{B_{n-1}} \mathrm{E}_t\left[u_{t+1}u'_{t+1}\right] \\
&\quad \times \left(\psi \frac{1}{1-h} F_c + \psi\left(1+\frac{1}{\phi}\right)\frac{\overline{n}}{1-\overline{n}} F_l + \psi F_\xi + F_\pi + (\gamma-\psi) F_j\right)'.
\end{aligned}$$

A.4 *Term premia*

We define returns under the expectations hypothesis using the assumption that future returns are discounted using the short-term rate, rather than the stochastic discount factor. Bond prices under the expectations hypothesis are therefore

$$B^{EH}_{t,n} = \frac{1}{I_t}\mathrm{E}_t\left[B^{EH}_{t+1,n-1}\right]$$

with 1-period yields defined as

$$\frac{1}{\left(I^{EH}_{t,n}\right)^n} = B^{EH}_{t,n}.$$

The term premium for maturity n, $TP_{t,n}$ is

$$TP_{t,n} = \frac{I_{t,n}}{I^{EH}_{t,n}}.$$

A.5 *Firms' optimization problem*

Under Rotemberg prices, firm j maximizes real profits

$$\max_{P^j_t} \mathrm{E}_t \sum_{s=t}^{\infty} Q_{t,s}\left[\frac{P^j_s Y^j_s}{P_s} - \frac{w_s}{P_s}\left(\frac{Y^j_s}{A_s}\right)^{\frac{1}{\alpha}} - \frac{\zeta}{2}\left(\frac{P^j_s}{P^j_{s-1}} - \left(\Pi^*_{s_\pi,s}\right)^{1-\iota}\Pi^{\iota}_{s-1}\right)^2 Y_s\right],$$

subject to the total demand for its output

$$Y_t(j) = \left(\frac{P_t(j)}{P_t}\right)^{-\theta} Y_t,$$

and to the production function

$$Y_t(j) = A_t L_t^{\alpha}(j),$$

where L_t is the labor index defined above.

The FOC is

$$\begin{aligned}
0 = {} & (1-\theta)\left(\frac{P_t^j}{P_t}\right)^{-\theta} Y_t \frac{1}{P_t} + \frac{\theta}{\alpha}\frac{w_t}{P_t}\left(\frac{Y_t}{A_t}\right)^{\frac{1}{\alpha}}\left(\frac{P_t^j}{P_t}\right)^{-\frac{\theta}{\alpha}-1}\frac{1}{P_t} \\
& - \zeta\left(\frac{P_t^j}{P_{t-1}^j} - \left(\Pi_{s_\pi,t}^*\right)^{1-\iota}\Pi_{t-1}^{\iota}\right) Y_t \frac{1}{P_{t-1}^j} \\
& + \mathrm{E}_t Q_{t,t+1}\zeta\left(\frac{P_{t+1}^j}{P_t^j} - \left(\Pi_{s_\pi,t+1}^*\right)^{1-\iota}\Pi_t^{\iota}\right) Y_{t+1}\frac{P_{t+1}^j}{P_t^j}\frac{1}{P_t^j},
\end{aligned}$$

or noting that all firms will set the same price and expressing variables in detrended form,

$$\begin{aligned}
& (\theta-1)\widetilde{Y}_t + \zeta\left(\Pi_t - \left(\Pi_{s_\pi,t}^*\right)^{1-\iota}\Pi_{t-1}^{\iota}\right)\widetilde{Y}_t\Pi_t \\
& = \frac{\theta}{\alpha}\frac{\widetilde{w}_t}{P_t}\frac{1}{Z_t^{\frac{1}{\alpha}}}\widetilde{Y}_t^{\frac{1}{\alpha}} + \mathrm{E}_t Q_{t,t+1}\zeta\left(\Pi_{t+1} - \left(\Pi_{s_\pi,t+1}^*\right)^{1-\iota}\Pi_t^{\iota}\right)\widetilde{Y}_{t+1}\Xi_{t+1}\Pi_{t+1}.
\end{aligned}$$

A.6 *Equilibrium*

Equilibrium is described by the following system:

- households

$$\begin{aligned}
\frac{\Lambda_t P_t}{u_{c,t}} &= (1-\beta)\widetilde{u}_t^{-\psi}\widetilde{J}_t^{\psi}, \\
-\frac{\widetilde{u}_{N,t}}{u_{c,t}} &= \frac{\theta_{w,t}-1}{\theta_{w,t}}\frac{\widetilde{w}_t}{P_t}, \\
\widetilde{J}_t^{1-\psi} &= (1-\beta)\widetilde{u}_t^{1-\psi} + \beta\left[\mathrm{E}_t\Xi_{t+1}^{1-\gamma}\widetilde{J}_{t+1}^{1-\gamma}\right]^{\frac{1-\psi}{1-\gamma}}, \\
\widetilde{u}_t &= u(\widetilde{C}_t - h\widetilde{C}_{t-1}, 1-N_t), \\
Q_{t,t+1} &= \beta\left[\mathrm{E}_t\widetilde{J}_{t+1}^{1-\gamma}\Xi_{t+1}^{1-\gamma}\right]^{\frac{\gamma-\psi}{1-\gamma}}\frac{\widetilde{J}_t^{\psi}}{\widetilde{J}_{t+1}^{\gamma}\Xi_{t+1}^{\gamma}}\frac{\Lambda_{t+1}}{\Lambda_t};
\end{aligned}$$

- firms

$$(\theta-1)\widetilde{Y}_t = -\zeta\big(\Pi_t - \big(\Pi^*_{s_{\pi,t}}\big)^{1-\iota}\Pi^{\iota}_{t-1}\big)\widetilde{Y}_t\Pi_t + \frac{\theta}{\alpha}\frac{\widetilde{w}_t}{P_t}\frac{1}{Z_t^{\frac{1}{\alpha}}}\widetilde{Y}_t^{\frac{1}{\alpha}}$$
$$+\,\mathrm{E}_t Q_{t,t+1}\zeta\big(\Pi_{t+1} - \big(\Pi^*_{s_{\pi,t+1}}\big)^{1-\iota}\Pi^{\iota}_{t}\big)\widetilde{Y}_{t+1}\Xi_{t+1}\Pi_{t+1};$$

- market clearing

$$\widetilde{Y}_t = \widetilde{C}_t + \widetilde{G}_t + \frac{\zeta}{2}\big(\Pi_t - \big(\Pi^*_{s_{\pi,t}}\big)^{1-\iota}\Pi^{\iota}_{t-1}\big)^2\widetilde{Y}_t,$$
$$N_t = \widetilde{Y}_t^{\frac{1}{\alpha}} Z_t^{-\frac{1}{\alpha}};$$

- policy rule

$$I_t = \left(\frac{\Pi^*_{s_{\pi,t}}\Xi_t^{\psi}}{\beta}\right)^{1-\rho_I}\left(\frac{\Pi_t}{\Pi^*_{s_{\pi,t}}}\right)^{\psi_\Pi}\left(\frac{\widetilde{Y}_t}{\widetilde{\overline{Y}}}\right)^{\psi_Y} I_{t-1}^{\rho_I} e^{\eta_{t+1}},$$
$$\ln\Pi^*_{s_{\pi,t}} = s_{\pi,t}\ln\Pi^*_{s_{\pi,H}} + (1-s_{\pi,t})\ln\Pi^*_{s_{\pi,L}};$$

- shocks

$$\Xi_t = \overline{\Xi}^{1-\rho_\xi}\Xi_{t-1}^{\rho_\xi}e^{\varepsilon_t^\xi}, \quad \varepsilon_{t+1}^\xi \sim N(0,\sigma_\xi),$$
$$\widetilde{G}_t = (g\widetilde{Y})^{1-\rho_g}\widetilde{G}_{t-1}^{\rho_g}e^{\varepsilon_t^g}, \quad \varepsilon_{t+1}^g \sim N(0,\sigma_g),$$
$$\mu_{w,t+1} = \mu_w^{1-\rho_\mu}(\mu_{w,t})^{\rho_\mu}e^{\varepsilon_{t+1}^\mu}, \quad \varepsilon_{t+1}^\mu \sim N(0,\sigma_\mu),$$
$$Z_t = Z_{t-1}^{\rho_z}e^{\varepsilon_t^z}, \quad \varepsilon_{t+1}^z \sim N(0,\sigma_{z,s_{z,t}}),$$
$$\eta_{t+1} = e^{\varepsilon_{t+1}^\eta}, \quad \varepsilon_{t+1}^\eta \sim N(0,\sigma_{\eta,s_{\eta,t}});$$

- standard deviations

$$\sigma_{z,s_{z,t}} = \sigma_{z,0}s_{z,t} + \sigma_{z,1}(1-s_{z,t}),$$
$$\sigma_{\eta,s_{\eta,t}} = \sigma_{\eta,0}s_{\eta,t} + \sigma_{\eta,1}(1-s_{\eta,t});$$

- C_{-1}, I_{-1}, Π_{-1} given.

A.7 *Numerical implementation*

For the numerical implementation of the model, we scale the maximum value function by a constant κ to increase accuracy. Define a dummy variable $\widetilde{D}_t = \mathrm{E}_t\Xi_{t+1}^{1-\gamma}\widetilde{J}_{t+1}^{1-\gamma}/\kappa^{1-\gamma}$. It follows that $\kappa^{1-\gamma}\widetilde{D}_t = \mathrm{E}_t\Xi_{t+1}^{1-\gamma}\widetilde{J}_{t+1}^{1-\gamma}$. This implies

$$\widetilde{D}_t = \frac{\mathrm{E}_t\Xi_{t+1}^{1-\gamma}\widetilde{J}_{t+1}^{1-\gamma}}{\kappa^{1-\gamma}},$$
$$\widetilde{J}_t^{1-\psi} = (1-\beta)\widetilde{u}_t^{1-\psi} + \beta\kappa^{1-\psi}\widetilde{D}_t^{\frac{1-\psi}{1-\gamma}},$$

$$Q_{t,t+1} = \beta\left(\frac{\kappa \widetilde{D}_t^{\frac{1}{1-\gamma}}}{\widetilde{J}_{t+1}}\right)^{\gamma-\psi}\left(\frac{\widetilde{u}_{t+1}}{\widetilde{u}_t}\right)^{-\psi}\frac{u_{c,t+1}}{u_{c,t}}\frac{1}{\Xi_{t+1}^{\gamma}}\frac{1}{\Pi_{t+1}}.$$

A.8 *Functional forms*

We rely on the Trabandt and Uhlig (2011) form for temporary utility, that is,

$$u_t = (C_t - h\Xi_t C_{t-1})\left(1-\eta(1-\psi)N_t^{1+\frac{1}{\phi}}\right)^{\frac{\psi}{1-\psi}}.$$

As a result,

$$\frac{\widetilde{w}_t}{P_t} = \frac{\eta\psi\left(1+\dfrac{1}{\phi}\right)(\widetilde{C}_t - h\widetilde{C}_{t-1})N_t^{\frac{1}{\phi}}}{1-\eta(1-\psi)N_t^{1+\frac{1}{\phi}}}\frac{\theta_{w,t}}{\theta_{w,t}-1},$$

$$\widetilde{J}_t^{1-\psi} = (1-\beta)(\widetilde{C}_t - h\widetilde{C}_{t-1})^{1-\psi}\left(1-\eta(1-\psi)N_t^{1+\frac{1}{\phi}}\right)^{\psi} + \beta\kappa^{1-\psi}\widetilde{D}_t^{\frac{1-\psi}{1-\gamma}},$$

$$Q_{t,t+1} = \beta\left(\frac{\kappa \widetilde{D}_t^{\frac{1}{1-\gamma}}}{\widetilde{J}_{t+1}}\right)^{\gamma-\psi}\left(\frac{\widetilde{C}_{t+1} - h\widetilde{C}_t}{\widetilde{C}_t - h\widetilde{C}_{t-1}}\right)^{-\psi}\left(\frac{1-\eta(1-\psi)N_{t+1}^{1+\frac{1}{\phi}}}{1-\eta(1-\psi)N_t^{1+\frac{1}{\phi}}}\right)^{\psi}\frac{1}{\Xi_{t+1}^{\gamma}}\frac{1}{\Pi_{t+1}},$$

$$(\theta-1)\widetilde{Y}_t = -\zeta\left(\Pi_t - \left(\Pi^*_{s_{\pi,t}}\right)^{1-\iota}\Pi_{t-1}^{\iota}\right)\widetilde{Y}_t\Pi_t + \frac{\theta}{\alpha}\frac{\widetilde{w}_t}{P_t}\left(\frac{\widetilde{Y}_t}{Z_t}\right)^{\frac{1}{\alpha}} + \cdots$$
$$+ \mathrm{E}_t Q_{t,t+1}\zeta\left(\Pi_{t+1} - \left(\Pi^*_{s_{\pi,t+1}}\right)^{1-\iota}\Pi_t^{\iota}\right)\widetilde{Y}_{t+1}\Xi_{t+1}\Pi_{t+1}.$$

A.9 *Elasticity of intertemporal substitution*

We compute the elasticity of intertemporal substitution of consumption as the elasticity of consumption to a change in the real interest rate holding labor supply constant.

Define the "consumption surplus" $\overleftrightarrow{c}_t \equiv \widetilde{C}_t - h\widetilde{C}_{t-1}$. The first-order approximation to the nominal stochastic discount factor

$$Q_{t,t+1} = \beta\left(\frac{\kappa \widetilde{D}_t^{\frac{1}{1-\gamma}}}{\widetilde{J}_{t+1}}\right)^{\gamma-\psi}\left(\frac{\overleftrightarrow{c}_{t+1}}{\overleftrightarrow{c}_t}\right)^{-\psi}\left(\frac{1-\eta(1-\psi)N_{t+1}^{1+\frac{1}{\phi}}}{1-\eta(1-\psi)N_t^{1+\frac{1}{\phi}}}\right)^{\psi}\frac{1}{\Xi_{t+1}^{\gamma}}\frac{1}{\Pi_{t+1}}$$

can be written as[S1]

$$\widehat{q}_{t,t+1} = -\psi\Delta\widehat{\overleftrightarrow{c}}_{t+1} - \psi\left(1+\frac{1}{\phi}\right)\frac{\overline{n}}{1-\overline{n}}\Delta\widehat{N}_{t+1} - \psi\widehat{\xi}_{t+1} - \widehat{\pi}_{t+1}$$
$$-(\gamma-\psi)\left(\widehat{\xi}_{t+1} + \widehat{\widetilde{j}}_{t+1} - \mathrm{E}_t[\widehat{\xi}_{t+1} + \widehat{\widetilde{j}}_{t+1}]\right),$$

where

$$\widehat{\widetilde{j}}_t + \widehat{\xi}_t = \sum_{i=0}^{\infty}(\beta\Xi^{1-\psi})^i\mathrm{E}_t\left[\widehat{\xi}_{t+i} + (1-\beta\Xi^{1-\psi})\left(\widehat{\overleftrightarrow{c}}_{t+i} - \frac{\psi}{1-\psi}\left(1+\frac{1}{\phi}\right)\frac{\overline{n}}{1-\overline{n}}\widehat{N}_{t+i}\right)\right].$$

[S1] In these derivations, $\kappa = 1$.

As a result,

$$\widehat{q}_{t,t+1} = -\psi \Delta \widehat{\overleftrightarrow{c}}_{t+1} - \psi \left(1 + \frac{1}{\phi}\right) \frac{\overline{n}}{1-\overline{n}} \Delta \widehat{N}_{t+1} - \psi \widehat{\xi}_{t+1} - \widehat{\pi}_{t+1},$$

and the real rate is

$$\widehat{r}_t = \psi \mathrm{E}_t \Delta \widehat{\overleftrightarrow{c}}_{t+1} + \psi \left(1 + \frac{1}{\phi}\right) \frac{\overline{n}}{1-\overline{n}} \mathrm{E}_t \Delta \widehat{N}_{t+1} + \psi \mathrm{E}_t \widehat{\xi}_{t+1}.$$

Rearranging terms

$$\widehat{\overleftrightarrow{c}}_t = -\frac{1}{\psi} \widehat{r}_t + \mathrm{E}_t \widehat{\overleftrightarrow{c}}_{t+1} + \frac{1}{\psi} \left(1 + \frac{1}{\phi}\right) \frac{\overline{n}}{1-\overline{n}} \mathrm{E}_t \Delta \widehat{N}_{t+1} + \mathrm{E}_t \widehat{\xi}_{t+1},$$

so that the long-run elasticity of substitution $\overline{EIS}$, that is, the elasticity which is obtained after households have adjusted their consumption habits, takes the usual value

$$\overline{EIS} = \frac{1}{\psi}.$$

Note that, in the absence of habits, this expression boils down to the usual value $1/\psi$.

To compute the short-run elasticity, we rewrite the consumption surplus in terms of the underlying consumption levels to obtain

$$\begin{aligned} \widehat{\widehat{c}}_t = & -\frac{1}{\psi} \frac{1-h}{1+h} \widehat{r}_t + \frac{1}{1+h} \mathrm{E}_t \widehat{\widehat{c}}_{t+1} + \frac{h}{1+h} \widehat{\widehat{c}}_{t-1} \\ & + \frac{1-h}{1+h} \left(1 + \frac{1}{\phi}\right) \frac{\overline{n}}{1-\overline{n}} \mathrm{E}_t \Delta \widehat{N}_{t+1} + \frac{1-h}{1+h} \mathrm{E}_t \widehat{\xi}_{t+1}. \end{aligned}$$

The short-run elasticity of substitution *EIS* is therefore

$$EIS = \frac{1}{\psi} \frac{1-h}{1+h},$$

which again boils down to $1/\psi$ when $h = 0$. Note that, since $h > 0$, it is always the case that $EIS < \overline{EIS}$.

A.10 *Model with equity prices*

In the version of the model in which we take equity prices into account, the household's budget constraint becomes

$$P_t C_t(j) + \mathrm{E}_t Q_{t,t+1} W_{t+1}(j) + \int_0^1 \mathrm{E}_t \Omega_{t+1}^j(i) P_t^S(i)\, \mathrm{d}i \leq \mathcal{W}_t(j) + (1+\tau_w) w_t(j) N_t(j) - T_t,$$

where total nominal household wealth $\mathcal{W}_t(j)$ is given by

$$\mathcal{W}_t(j) = W_t(j) + \int_0^1 \Omega_t^j(i) \big(P_t^S(i) + P_t \delta_t(i)\big)\, \mathrm{d}i,$$

that is, it includes two components. The first component, as before, is a complete portfolio of state-contingent bonds $W_t(j)$ yielding a return $\mathrm{E}_t Q_{t,t+1} W_{t+1}(j)$ in $t+1$. The second component are cum-dividend shares in each firm in the economy, where $\Omega_t^j(i)$ is the household's portfolio holding of shares in firm i, $P_t^S(i)$ is the share price, and $\delta_t(i)$ is the dividend.

The first-order conditions will now include the choice of the shares in each firm i,

$$U_{v,t}\left[\mathrm{E}_t J^{1-\gamma}(V_{t+1})\right]^{\frac{\gamma}{1-\gamma}} J^{-\gamma}(V_{t+1}) J_V(V_{t+1})\left(P_{t+1}^S(i) + P_{t+1}\delta_{t+1}(i)\right) = \Lambda_t P_t^S(i)$$

or, using the same definitions as above

$$\frac{P_t^S(i)}{P_t} = Q_{t,t+1}\left(\frac{P_{t+1}^S(i)}{P_{t+1}} + \delta_{t+1}(i)\right)\Pi_{t+1}.$$

Aggregate real dividend payments and aggregate real market capitalization can be defined as

$$P_t^S = \int_0^1 P_t^S(i)\,\mathrm{d}i,$$
$$\delta_t = \int_0^1 \delta_t(i)\,\mathrm{d}i$$

so that the detrended, real stock market value $\widetilde{p}_t^S = p_t^S/B_t$ (where $p_t^S \equiv P_t^S/P_t$ and $\widetilde{\delta}_t = \delta_t/B_t$) is given by

$$\widetilde{p}_t^S = Q_{t,t+1}\Xi_{t+1}\left(\widetilde{p}_{t+1}^S + \widetilde{\delta}_{t+1}\right)\Pi_{t+1}.$$

Equilibrium dividends can be obtained from households' total income

$$\widetilde{\delta}_t = \widetilde{Y}_t - (1+\tau_w)\frac{\widetilde{w}_t}{P_t}\left(\frac{\widetilde{Y}_t}{Z_t}\right)^{\frac{1}{\alpha}}.$$

The ex ante nominal return on equity $I_{e,t} = \mathrm{E}_t \frac{p_{t+1}^S + \delta_{t+1}}{p_t^S}\Pi_{t+1}$ can be rewritten in terms of detrended variables as

$$I_{e,t} = \mathrm{E}_t \frac{\widetilde{p}_{t+1}^S + \tilde{\delta}_{t+1}}{\widetilde{p}_t^S}\Xi_{t+1}\Pi_{t+1}.$$

We define the equity premium in terms of nominal returns as $ERP_t^n = I_{e,t}/I_t$ or

$$ERP_t^n = \frac{1}{I_t}\mathrm{E}_t \frac{\widetilde{p}_{t+1}^S + \tilde{\delta}_{t+1}}{\widetilde{p}_t^S}\Xi_{t+1}\Pi_{t+1}.$$

Note that in the nonstochastic steady state

$$ERP^n = 0.$$

In the impulse responses, we also use the ex ante real interest rate $\bar{R}_t$ defined as

$$\bar{R}_t = \frac{I_t}{\mathrm{E}_t \Pi_{t+1}}.$$

Appendix B: Model estimation

B.1 *Data definition and sources*

The macroeconomic data series we use for estimation are obtained from FRED [S2] and their mnemonics and original sources are as follows:

- Civilian Non-inst. Population: `CNP16OV`;
- Real GDP: `GDPC1`;
- Real PCE: `PCECC96`;
- Nominal PCE: `PCEC`;
- Effective Federal Funds Rate: `FFR`.

The data on continuously compounded yields on 3-year (`SVENY03`) and 10-year (`SVENY10`) zero-coupon bonds come from Gürkaynak, Sack, and Wright (2007).[S3]

B.2 *Prior specification details*

For the ϕ parameter, we rely on a normal prior centred around 1.0, a value in between macro estimates and micro estimates of the Frisch elasticity of labor supply (see, e.g., the evidence reviewed in Chetty, Guren, Manoli, and Weber (2011)). We use a shifted Gamma distribution for ψ and γ, to ensure that ψ, $\gamma > 1$. We center the distribution of ψ around a value above but close to 1. For the γ parameter, which contributes to shape risk aversion, we use a very large standard deviation whose prior 95% confidence set goes from 2 to 30. The habit parameter has a beta prior centered around 0.5. Finally, for β we use a relatively tight prior with a mean of 0.9985. This is consistent with assumptions made in models with growth; see, for example, Christiano, Motto, and Rostagno (2014).

For the long run parameter Ξ, we rely on a more dogmatic prior. We use a tight prior centred around an annualized value of 2%, which is consistent with the average per-capita U.S. GDP/GNP growth from the 1870s to the 1950s; see Maddison (2003).

The price adjustment cost ζ is typically calibrated based on the implied frequency of adjustment of prices in linearized models. In our model, however, the relationship is more complex due to both the nonlinearity of the model and the presence of steady-state inflation. We therefore center the prior around 15, which is roughly consistent, for example, with the value used in Schmitt-Grohé and Uribe (2004), but allow for a relatively large standard deviation. For inflation indexation, we rely on a beta prior centered around 0.5.

[S2] https://fred.stlouisfed.org/

[S3] https://www.federalreserve.gov/data/yield-curve-tables/feds200628.csv

The elasticity of intratemporal substitution θ, which is weakly identified, is set dogmatically at 6. Similarly, we set the steady-state gross wage mark-up μ_w to 1.2.

B.3 *Likelihood computation*

Solving the model to second order, we obtain the reduced-form system of equations,

$$y^o_{t+1} = k_{y,j} + F\hat{x}_{t+1} + \frac{1}{2}\big(I_{n_y} \otimes \hat{x}'_{t+1}\big)E\hat{x}_{t+1} + Dv_{t+1}, \tag{S1}$$

$$\hat{x}_{t+1} = k_{x,i} + P\hat{x}_t + \frac{1}{2}\big(I_{n_x} \otimes \hat{x}'_t\big)G\hat{x}_t + \widetilde{\sigma}\Sigma_i w_{t+1}, \tag{S2}$$

$$s_t \backsim MS(Q), \tag{S3}$$

where

$$\begin{aligned} k_{y,j} &= k_{y,s_{t+1}=j}, \\ k_{x,i} &= k_{x,s_t=i}, \\ \Sigma_i &= \Sigma(s_t = i). \end{aligned}$$

and Q is the transition probability matrix associated with the Markov switching (MS) process s_t.

The vector y^o_t includes all observable variables, and v_{t+1} and w_{t+1} are measurement and structural shocks, respectively. In this representation, as shown in Amisano and Tristani (2011), the regime switching variables affect the system by changing the intercepts $k_{y,j}$, $k_{x,i}$, and the loadings of the structural innovations Σ_i (we indicate here with i the value of the discrete state variables at t and with j the value of the discrete state variables at $t+1$).

To compute the approximate likelihood, at any point in time we first linearize the two quadratic terms around the conditional mean of the continuous state variables conditional on the prevailing regime. As a result, the two equations above can be rewritten as

$$\begin{aligned} y^o_{t+1} &= \widetilde{k}^{(i,j)}_{y,t+1} + \widetilde{F}^{(i,j)}_{t+1}\hat{x}_{t+1} + Dv_{t+1}, \\ \widehat{x}_{t+1} &= \widetilde{k}^{(i)}_{x,t} + \widetilde{P}^{(i)}_t\widehat{x}_t + \Sigma_i w_{t+1}, \end{aligned}$$

where

$$\begin{aligned} \widetilde{k}^{(i,j)}_{y,t+1} &= \widetilde{k}_{y,j} + \frac{1}{2}\big(I_{n_y} \otimes \hat{x}^{(i)'}_{t+1|t}\big)E\hat{x}^{(i)}_{t+1|t} - \Delta_{i,t+1}\hat{x}^{(i)}_{t+1|t}, \\ \widetilde{F}^{(i,j)}_{t+1} &= F + \Delta_{i,t+1}\hat{x}^{(i)}_{t+1|t} = E\big(x_{t+1}|\underline{y}^o_{1:t}, s_t = i, \theta\big), \\ \Delta_{i,t+1} &= \left[\frac{\partial\left(\frac{1}{2}\big(I_{n_y} \otimes \hat{x}'_{t+1}\big)E\hat{x}_{t+1}\right)}{\partial\hat{x}_{t+1}}\right]_{\hat{x}_{t+1}=\hat{x}^{(i)}_{t+1|t}}, \end{aligned}$$

$$\widetilde{k}_{x,t}^{(i)} = \widetilde{k}_{x,i} + \frac{1}{2}\big(I_{n_x} \otimes \hat{x}_{t|t}^{(i)\prime}\big)G\hat{x}_{t|t}^{(i)} - \Delta_{i,t}\hat{x}_{t|t}^{(i)},$$

$$\widetilde{P}_t^{(i)} = P + \Delta_{i,t}\hat{x}_{t|t}^{(i)} = E\big(\hat{x}_t|\underline{y}_{1:t}^o, s_t = i, \theta\big),$$

$$\boldsymbol{\Delta}_{i,t} = \left[\frac{\partial\left(\frac{1}{2}\big(I_{n_x} \otimes \hat{x}_t'\big)G\hat{x}_t\right)}{\partial \hat{x}_t}\right]_{\hat{x}_t = \hat{x}_{t|t}^{(i)}}$$

for regime-dependent intercepts $\widetilde{k}_{y,t+1}^{(i,j)}$, $\widetilde{k}_{x,t}^{(i)}$ and slope coefficients $\widetilde{F}_{t+1}^{(i,j)}$, $\widetilde{P}_t^{(i)}$. We then apply Kim's (1994) approximate filter to forecast

$$\hat{x}_{t+1|t}^{(i,j)} = \widetilde{k}_{x,t}^{(i)} + \widetilde{P}_t^{(i)}\hat{x}_{t|t}^{(i)} = \hat{x}_{t+1|t}^{(i)},$$

$$Q_{t+1|t}^{(i,j)} = \widetilde{P}_t^{(i)}Q_{t|t}^{(i,j)}\widetilde{P}_t^{(i)\prime} + \Sigma_i\Sigma_i' = Q_{t+1|t}^{(i)},$$

and update the vector of continuous state variables

$$\hat{x}_{t+1|t+1}^{(j)} = \sum_{i=1}^{m}\hat{x}_{t+1|t+1}^{(i,j)} \times p(s_t = i|s_{t+1} = j, \underline{y}_{1:t+1}),$$

$$Q_{t+1|t+1}^{(j)} = \sum_{i=1}^{m}\Big[\big(\hat{x}_{t+1|t+1}^{(i,j)} - \hat{x}_{t+1|t+1}^{(j)}\big)\big(\hat{x}_{t+1|t+1}^{(i,j)} - \hat{x}_{t+1|t+1}^{(j)}\big)' + Q_{t+1|t+1}^{(i,j)}\Big] \times p(s_t = i|s_{t+1} = j, \underline{y}_{1:t+1}),$$

and then update the regime probabilities

$$p(s_{t+1} = j, s_t = i|\underline{y}_{1:t}) = p_{ij,t+1|t} = p_{ij} \times p(s_t = i|\underline{y}_{1:t}),$$

and

$$p(s_{t+1} = j, s_t = i|\underline{y}_{t+1}) = p_{ij,t+1|t} \times \frac{p(y_{t+1}|\underline{y}_t, s_{t+1} = j, s_t = i)}{p(y_{t+1}|\underline{y}_t)},$$

$$p(s_{t+1} = j|\underline{y}_{1:t+1}) = \sum_{i=1}^{m} p(s_{t+1} = j, s_t = i|\underline{y}_{1:t+1}),$$

$$p(s_t = i|s_{t+1} = j, \underline{y}_{1:t+1}) = \frac{p(s_{t+1} = j, s_t = i|\underline{y}_{1:t+1})}{p(s_{t+1} = j|\underline{y}_{1:t+1})},$$

$$p(y_{t+1}|\underline{y}_{1:t}) = \sum_{i=1}^{m}\sum_{j=1}^{m} p(y_{t+1}|\underline{y}_{1:t}, s_{t+1} = j, s_t = i) \times p(s_{t+1} = j, s_t = i|\underline{y}_{1:t}).$$

The conditional log-likelihood is obtained as

$$\log L = \sum_{t=1}^{T}\log p(y_{t+1}|y_{1:t}).$$

B.3.1 *Accuracy of likelihood computations* As mentioned in Section 4.3 of the paper, the version of the model that receives most support from the data is the one in which the standard deviations of measurement errors are estimated (the `baseline` version in Table 2 in the paper). It turns out that the estimated values of these parameters are too small to be compatible with likelihood computations via Sequential Monte Carlo (SMC). For this reason, we decided to compute the likelihood by using the approximation based on the Extended Kalman Filter and the Kim filter (EKF + Kim) just described above. In order to informally measure the properties of our EKF + Kim approximation, we consider the model with calibrated measurement errors standard deviations, for which likelihood evaluation by SMC is possible, and we use $\hat{\theta}$, the posterior mean of the parameters estimated with the EKF + Kim approximation, as a reference to compare the likelihood computations based on SMC and on our EKF + Kim approximation.

In order to describe the SMC algorithm that we use to compute the likelihood, let us define:

- $x_t^{(i)}$, $i = 1, 2, \ldots, N$, a swarm of particles each drawn from each distribution conditional on data and observations up to time t ($y_{1:t}$), that is, $p(x_t|y_{1:t}, \hat{\theta})$. Note that x_t denotes the full vector of state variables, discrete and continuous.

- $p(x_t|x_{t-1}, \hat{\theta})$, the conditional state density implied by the state equations using $\hat{\theta}$ as parameters.

- $p(y_t|x_t, \hat{\theta})$,the conditional density of observable variables implied by the measurement equations.

- $\tilde{p}(x_t|x_{t-1}, y_{1:t}, \hat{\theta}) = \tilde{p}(x_t|x_{t-1}, y_t, \hat{\theta})$ the state density obtained running the EKF while conditioning on x_t.

The SMC algorithm works as follows:

- draw $x_0^{(i)}$, $i = 1, 2, \ldots, N$, from its ergodic distribution;

- for each $t = 1, 2, \ldots, T$,
  - draw $x_t^{(i)}$, $i = 1, 2, \ldots, N$ from $\tilde{p}(x_t|x_{t-1}^{(i)}, y_{1:t}, \hat{\theta})$;
  - assign each particle the weight $w_t^{(i)} = \frac{p(x_t^{(i)}|x_{t-1}^{(i)}, \hat{\theta}) \times p(y_t|x_t^{(i)}, \hat{\theta})}{\tilde{p}(x_t^{(i)}|x_{t-1}^{(i)}, y_t, \hat{\theta})}$ and resample the particles using those weights;

At each point in time, the sample mean (across particles) of the weights is a consistent estimate of $p(y_t|y_{t-1}, \hat{\theta})$, the conditional likelihood of each observation; the product of these terms over the sample is the SMC likelihood evaluation. At each point in time, the sample mean (across particles) of the terms $\tilde{p}(y_t|x_t - 1^{(i)}, \hat{\theta})$, that is, the conditional density of observable variables obtained conditioning on the state at $t - 1$ and using the EKF, provides a consistent estimate of the conditional likelihood that does not rely on the Kim filter approximation, but still uses the stepwise EKF linearization. We call the resulting likelihood "EKF + SMC."

TABLE B.1. Likelihood computations comparisons: correlations between Kim + EKF, EKF + SMC, full SMC methods.

	Kim + EKF	EKF + SMC
EKF + SMC	0.98	–
SMC	0.96	0.98

By comparing the SMC, EKF + SMC, and EKF + Kim likelihood computations, we can parse the effects of each of the two approximations that we use for likelihood computations (that is using the Kim algorithm with the EKF): the EKF + SMC approach eliminates recourse to the Kim's approximation, while the SMC approach eschews any kind of approximation. Since the SMC likelihood is computed here only for one set of parameter values, we could splurge on the number of particles ($N = 500{,}000$) and get a highly accurate measure of the conditional likelihood $p(y_t|y_{1:t-1}, \hat{\theta})$ for each observation in the sample.

Table B.1 reports the correlation among log likelihood computations throughout the sample (i.e., $\log(p(y_t|y_{t-1}, \hat{\theta}))$, $t = 1, 2, \ldots, T$) across the three methods, and shows very high, although not perfect correlations. In particular, as Figure B.1 shows, more notable differences across the methods occur in correspondence with outliers (when the log likelihood dips down). It is interesting to note that outliers correspond to particularly large shocks that takes the model away from its stochastic steady state. In these circumstances, the quality of the local approximation used to solve the model tends to deteriorate as well.

Since the results of our paper are mainly in terms of the model implications for the latent variables, we thought it would be interesting to compare the filtered latent variables that we get with our approach with those obtained by using the SMC approach. As

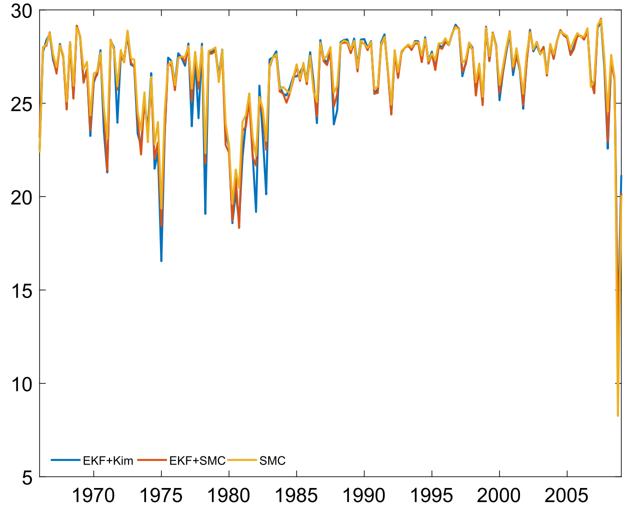

FIGURE B.1. Likelihood computations: EKF + Kim, EKF + SMC, SMC.

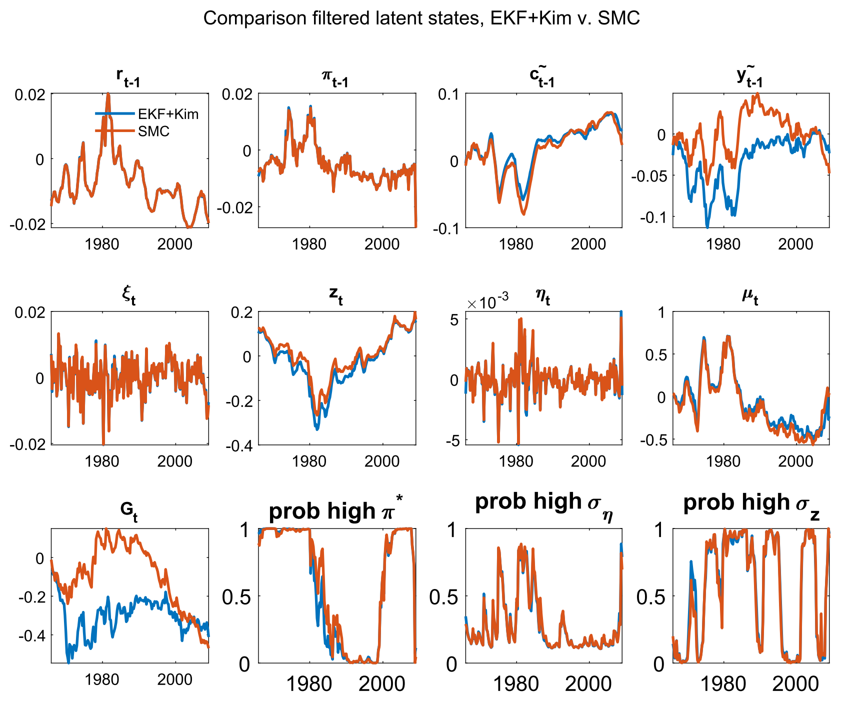

FIGURE B.2. Latent variables computations: EKF + Kim versus SMC.

Figure B.2 shows the filtered latent variables are very similar across the two approaches, in most cases on top of each other. There are a couple of exceptions, namely the lagged detrended "true" output (*lag_y*, net of measurement error) and government spending (G). These differences stem from the combined effect of large shocks on G and large measurement errors on output which, in turn produce high filtering uncertainty on both variables.

B.4 *MCMC simulation*

We start by computing the mode of the posterior distribution of the parameters by using a two-step approach:

1. we compute a reasonable approximation to the mode by using a simulated annealing algorithm (Goffe, Ferrier, and Rogers (1994));

2. using the result from the first step as initial value, we then run a gradient based method (C. Sims's `csminwell`) to find the posterior mode.

Having found the posterior mode, we compute the Hessian of the log posterior distribution at the mode and we use minus the inverse of this matrix as covariance matrix for a Gaussian distribution in a random walk Metropolis–Hastings algorithm, as customarily done in Bayesian estimation of DSGE models (as described in An and Schorfheide (2007)). This covariance matrix is scaled to achieve acceptance rates of 50%.

The MCMC algorithm is run to obtain 440,000 draws, the first 40,000 are discarded and the remaining ones are thinned (i.e., one every 20 draws is recorded), resulting in a final posterior sample of 20,000 draws, which is then used in all the computations reported in the paper.

We find that the resulting posterior sample has good properties in terms of acceptance rate and low correlation across draws.

B.5 *Unconditional and conditional moments*

The computation of first and second unconditional and conditional moments entails different levels of difficulty, depending on whether the Markov switching processes affect shock variances only (like in the version of the model with a constant inflation target), or they do affect parameters that enter in the model's steady state (such as the model with different regimes on the inflation target). We provide a separate description of the two cases.

B.5.1 *Regime switches affecting shock variances only* As we have shown in Section 3.2 of the paper, in this case the quadratic approximation of the model's solution has regime-specific intercepts and loading matrices for the shocks, while the linear slope and quadratic terms are constant. This feature of the solution allows us to compute first- and second-order moments analytically via a pruning approach, that is, we take into consideration only linear terms for the computation of second-order moments, and linear and quadratic terms for the computation of first-order moments. The computation of unconditional moments works as follows: for each draw of the parameter vector from the posterior distribution, we compute the state space representation (S1), (S2), and (S3). From the state space representation, we obtain the unconditional covariance matrix of state vector shocks as

$$\Omega_{ww} = \widetilde{\sigma} \sum_{i=1}^{m} \Sigma_i \Sigma_i' \pi_i,$$

where π_i are the ergodic state probabilities associated with the transition probability matrix Q. Taking the state equation stripped of its second-order term, we can obtain $\Omega_{xx,0}$, the static covariance matrix of $\widehat{x}_t$, as a solution of

$$\mathrm{Cov}(\widehat{x}_t) = \Omega_{xx,0} = P\Omega_{xx,0}P' + \Omega_{ww}.$$

Dynamic covariance matrices are obtained by applying the recursion

$$\mathrm{Cov}(\widehat{x}_t, \widehat{x}_{t-j}) = \Omega_{xx,j} = P\Omega_{xx,j-1}, \quad j = 1, 2, 3, \ldots$$

and the covariance matrices for the variables in y_t^o are obtained using the corresponding linear measurement equation:

$$\Omega_{yy,0} = \mathrm{Cov}\big(y_t^o\big) = F\Omega_{xx,0}F' + DD',$$

$$\Omega_{yy,j} = \mathrm{Cov}\big(y_t^o, y_{t-j}^o\big) = F\Omega_{xx,j}F', \quad j = 1, 2, 3, \ldots.$$

When computing first moments, we take into consideration both first and second order. To show how first moments are obtained, we rewrite the state space representation in equivalent form as

$$y^o_{t+1} = k_{y,j} + F\hat{x}_{t+1} + \frac{1}{2}\overline{E}\,\mathrm{vec}(\hat{x}_{t+1}\hat{x}'_{t+1}) + Dv_{t+1}, \tag{S4}$$

$$\hat{x}_{t+1} = k_{x,i} + P\hat{x}_t + \frac{1}{2}\overline{G}\,\mathrm{vec}(\hat{x}_t\hat{x}'_t) + \widetilde{\sigma}\Sigma_i w_{t+1}, \tag{S5}$$

where $\overline{E}$ and $\overline{G}$ are obtained by suitably rearranging the elements of the matrices E and G, respectively. Taking the unconditional expected value of the two expressions above yields the first moments:

$$\mu_y = E(y^o_t) = k_y + F\mu_x + \frac{1}{2}\overline{E}\,\mathrm{vec}(\Omega_{xx,0}), \quad k_y = \sum_{j=1}^{m} k_{y,j}\pi_j,$$

$$\mu_x = E(\widehat{x}_t) = [I_{n_x} - P]^{-1}\left[k_x + \frac{1}{2}\overline{G}\,\mathrm{vec}(\Omega_{xx,0})\right], \quad k_x = \sum_{i=1}^{m} k_{x,i}\pi_i.$$

The computation of conditional first- and second-order moment is straightforward.

B.5.2 *Regime switching affecting the nonstochastic steady state* As we have described in Section 3.2 of the paper, in this case the second-order approximation is obtained by including the regime-switching parameter in the state vector and applying perturbation around the state vector's ergodic mean. The resulting solution can be rewritten exclusively in terms of the continuous states, but the linear slope terms become regime-specific. In this case, the computations of theoreical moments is more difficult. In principle, one could apply a version of the approach used in Bianchi (2016) combined with pruning. Alternatively, (and this is the approach that we have used in this version of the paper), we have computed unconditional and conditional moments by suitably simulating a large number of paths for the observable variables subject to pruning and computing the required moments across the different simulations. This approach is computationally demanding but straightforward to be implemented.

B.6 *Impulse response functions*

To compute impulse response functions (IRFs), we follow Koop, Pesaran, and Potter (1996). IRFs can be computed with respect to all shocks hitting the model, either continuous (the shocks in the state vector w_t) or discrete, that is, the shocks that lead to a change in the discrete Markov switching process that affects the model. We define ε_t as the vector containing all the shocks affecting continuous and discrete states. We compute IRFs to a shock ε_{jt} of size δ_j occurring at time t,using the following algorithm:

- draw $\theta^{(i)}$, $i = 1, 2, \dots, M$, from the posterior distribution of the parameters;
- compute the state space representation corresponding to $\theta^{(i)}$, run the Kalman filter and draw $\widehat{x}^{(i)}_t$, $s^{(i)}_t$ from their joint posterior distribution conditional on $\theta^{(i)}$;

- draw two histories of shocks $\varepsilon_{t+h}^{(i,1)} \varepsilon_{t+h}^{(i,2)}$, $h = 0, 1, 2, \ldots, H$, which are totally identical but differ only for the shock ε_{jt}, such that

$$\varepsilon_{jt}^{(i,2)} = \varepsilon_{jt}^{(i,1)} + \delta_j;$$

- feed these two histories of shocks to state and measurement equations starting from $\widehat{x}_t^{(i)}$, $s_t^{(i)}$, and generate 2 paths

$$y_{t+h}^{(i,1)}, y_{t+h}^{(i,2)}, \quad h = 0, 1, 2, \ldots, H;$$

  the difference between these two path traces the dynamic response of shock δ_j;

- the empirical distribution of this difference across draws $\theta^{(i)}$ gives the posterior distribution of the IRFs.

Note that IRFs reported in Figure 7 in the paper are obtained by fixing the state s_{t+h}, $h = 0, 1, 2, \ldots, H$ at the value corresponding to low volatility for all the shocks. IRFs reported in Figure 5 are obtained by contemplating a one-off shift in volatility, that is, forcing the process to move to the high volatility state only once at time t.

B.7 *Variance decomposition*

Forecast Error Variance decomposition (FEVD) is a measure of the importance of the model's orthogonal shocks in determining the observed behavior of each variable in the model at different horizons. Tables B.2, B.3, B.4 report the variance decomposition for selected model variables at 1, 12, 40 horizons, computed using the posterior mean of the parameter values. The tables show, among other things, how important is the role of switches in the technology shock variance on long term rates and the equity premium at all horizons.

The procedure to compute variance decomposition is straightforward in linear models and a bit more complicated in nonlinear ones, such as the quadratic MS-DSGE model used in our paper. In particular, difficulties arise since:

TABLE B.2. Variance decomposition 1 step ahead.

	ξ	z	*pol*	μ	G	s_{π^*}	s_{pol}	s_z	*meas.err.*
π_t	5.8	12.5	6.3	36.0	9.8	22.1	4.4	3.2	0.0
Δc	58.3	0.0	0.9	1.0	0.3	0.3	0.6	0.6	38.0
Δy	37.7	0.0	0.2	1.3	26.9	0.1	0.3	0.2	33.3
r	20.8	5.2	24.7	15.1	8.2	8.1	16.1	1.7	0.0
$Rbar_t$	20.7	7.7	24.3	7.9	4.1	16.3	16.5	2.4	0.0
R_n	0.0	92.2	0.0	0.0	4.4	0.0	0.0	3.3	0.0
$EHR40$	0.2	35.6	0.1	0.7	10.3	41.2	0.0	11.9	0.0
$R40$	0.1	37.3	0.0	0.6	8.2	33.2	0.0	0.2	20.3
$ERPn_t$	0.0	0.0	0.0	0.0	0.0	0.0	0.7	99.3	0.0
PiEqbar_real	0.3	72.1	5.2	5.3	6.3	4.5	3.4	2.9	0.0

Table B.3. Variance decomposition 12 steps ahead.

	ξ	z	pol	μ	G	s_{π^*}	s_{pol}	s_z	$meas.err.$
π_t	5.5	9.0	6.2	26.3	8.4	18.0	12.6	14.0	0.0
Δc	53.0	1.7	0.5	6.6	0.5	1.2	1.6	1.8	33.2
Δy	38.2	1.2	0.6	7.4	19.2	1.6	1.6	1.3	28.8
r	3.7	15.5	5.0	13.3	15.6	16.2	10.0	20.6	0.0
$Rbar_t$	12.3	5.2	15.2	18.7	5.6	4.2	30.9	7.8	0.0
R_n	0.0	58.6	0.0	0.0	3.8	0.0	0.0	37.6	0.0
$EHR40$	0.0	42.5	0.1	0.7	6.1	20.3	0.2	30.1	0.0
$R40$	0.1	49.7	0.1	0.7	6.0	19.3	0.2	21.2	2.8
$ERPn_t$	0.0	0.0	0.0	0.0	0.0	0.0	0.4	99.6	0.0
$PiEqbar\_real$	1.1	44.5	4.6	4.3	3.3	1.5	9.4	31.3	0.0

1. the model is nonlinear;
2. there are shocks in the variances, that is, discrete shocks, beside the usual continuous shocks;
3. there is uncertainty around the latent states, even conditioning on parameter values.

It is important to notice though, that the nonlinearities generated by quadratic terms in the model's solution do not play any role if second-order moments are computed using an appropriate pruning procedure, that is, taking into consideration only linear terms. In order to describe how the variance decomposition results contained in the paper are computed, we define

$$v^{(i)}(j, h, \{S\}) = V\left(y_{it+h}|y_{1:t}, \theta^{(i)}_{\{S\}}\right)$$

Table B.4. Variance decomposition 40 steps ahead.

	ξ	z	pol	μ	G	s_{π^*}	s_{pol}	s_z	$meas.err.$
π_t	4.9	9.2	6.0	23.0	7.4	17.1	15.1	17.2	0.0
Δc	50.0	0.6	1.2	7.4	0.6	1.4	2.6	2.4	33.7
Δy	35.3	1.4	0.8	6.6	22.0	0.9	2.6	2.0	28.3
r	3.0	17.4	2.5	10.3	11.6	13.7	7.1	34.5	0.0
$Rbar_t$	11.4	3.7	12.1	16.9	6.0	5.2	32.4	12.4	0.0
R_n	0.0	47.2	0.0	0.0	2.6	0.0	0.0	50.1	0.0
$EHR40$	0.1	44.3	0.1	0.7	2.7	7.9	0.2	44.2	0.0
$R40$	0.1	47.4	0.1	0.7	2.7	7.7	0.2	40.2	1.0
$ERPn_t$	0.0	0.0	0.0	0.0	0.0	0.0	0.1	99.9	0.0
$PiEqbar\_real$	0.6	40.1	3.3	2.9	3.5	0.9	7.4	41.3	0.0

Note: π, Δy, i, $Rbar$, R_n, $EHR40$, $R40$, ERP, $PiEqbar\_real$ denote, respectively, inflation, output growth, the short term interest rate, the short real rate, the natural rate of interest, the long-run (40-period) rate compatible with the expectation hypothesis, the actual long-run rate, the equity premium, and the deflated and detrended equity price.

the conditional variance of y_{it+h} conditional on the parameter vector $\theta^{(i)}$ drawn from the joint posterior distribution, and $\{S\}$ denotes the set of shocks or sources of randomness being allowed to be active in the system from $t+1$ to $t+H$, the end of the projection period.

As an example, setting $S=\{0\}$ means that all all shocks in the system (continuous and discrete) are being switched off, and this is achieved by using $\theta^{(i)}_{\{0\}}$, a modified version of $\theta^{(i)}$ to $\theta^{(i)}_{\{0\}}$ in which the standard deviations of all shocks have been set to zero. In this case, the conditional variance is determined only by the conditional variance of all the latent variables (continuous and discrete) at time t, what is usually referred to as "initial condition." When we instead define S to be the full set of shocks, we compute conditional variances using $\theta^{(i)}$. These variances are determined by the full structure of shocks in the model.

In order to describe the portion of forecast variances attributable to each shock, let us call ε^k_t, $k=1,2,3,4,5$, the continuous shocks in the model, the first three of them having Markov switching variances.

FEVD coefficients are computed as follows:

- for each value of the parameters $\theta^{(i)}$, $i=1,2,\dots,M$, drawn from the posterior distribution, we compute the solution for the model and the theoretical h-step ahead forecast variances of all observed series. This is done using the appropriate pruning, that is, considering only linear terms. These conditional variances, generated when all shocks are active, enter in the denominator of the FEVD coefficients. This is the denominator of any FEVD coefficient and it is indicated as

$$v^{(i)}\big(j,h,\{all\}\big)=V\big(y_{it+h}|y_{1:t},\theta^{(i)}\big).$$

- Starting from $\theta^{(i)}$, we set the standard deviations of all shocks (and measurement errors) to zero, and we obtain $\theta^{(i)}_{\{0\}}$. We then compute the associated forecast variances. This is the portion of variances due to uncertainty around initial conditions:

$$v^{(i)}\big(j,h,\{0\}\big)=V\big(y_{it+h}|y_{1:t},\theta^{(i)}_{\{0\}}\big),$$

$$\mathrm{FEVD}^{(i)}\big(j,h,\{0\}\big)=100\times\frac{v^{(i)}\big(j,h,\{0\}\big)}{v^{(i)}\big(j,h,\{all\}\big)}.$$

- Starting from $\theta^{(i)}_{\{0\}}$, for each of the continuous shocks with Markov switching variances (ε^k_t, $k=1,2,3$), we first consider the contribution of the shock by setting its two variances both equal to its low volatility regime value, that is, $\sigma_{1,k}$ $k=1,2,3$, therefore obtaining the vector $\theta^{(i)}_{\{0,k\}}$. In this way, we introduce only the k^{th} continuous shock, but we zero out its variance jumps. We indicate the corresponding variance as

$$v^{(i)}\big(j,h,\{0,k\}\big)=V\big(y_{it+h}|y_{1:t},\theta^{(i)}_{\{0,k\}}\big),$$

  and we isolate the contribution of that shock by netting out the effect of the initial condition as follows:

$$v^{(i)}\big(j,h,\{k\}\big)=v^{(i)}\big(j,h,\{0,k\}\big)-v^{(i)}\big(j,h,\{k\}\big)$$

and

$$\text{FEVD}^{(i)}\big(j, h, \{k\}\big) = 100 \times \frac{v^{(i)}\big(j, h, \{k\}\big)}{v^{(i)}\big(j, h, \{all\}\big)}.$$

- For each of the shocks with switching variances, we define $\theta^{(i)}_{\{0,k,s_k\}}$ the modification of $\theta^{(i)}_{\{0,k\}}$ where the two variances of shock ε^k_t are set to their respective high and low values. In this way, that shock is allowed to be heteroskedastic. The corresponding conditional variances are then

$$v^{(i)}\big(j, h, \{0, k, s_k\}\big) = V\big(y_{it+h}|y_{1:t}, \theta^{(i)}_{\{0,k,s_k\}}\big),$$

  and we isolate the contribution of the k^{th} shock variance jumps by subtracting the portion of variance jointly due to the initial condition and to the k^{th} shock when assumed to be homoskedastic:

$$v^{(i)}\big(j, h, \{s_k\}\big) = v^{(i)}\big(j, h, \{0, k, s_k\}\big) - v^{(i)}\big(j, h, \{0, k\}\big);$$

  the FEVD of the k^{th} shock Markov switching jumps is hence computed as follows:

$$\text{FEVD}^{(i)}\big(j, h, \{s_k\}\big) = 100 \times \frac{v^{(i)}\big(j, h, \{s_k\}\big)}{v^{(i)}\big(j, h, \{all\}\big)}.$$

- For each of the 8 continuous shocks without Markov-switching variance, that is, the mark-up shock ε^{μ}_t, the permanent technology shock ε^{ξ}_t, and the 6 measurement errors ($l = 1, 2, \ldots, 8$), we measure the FEVD contribution by defining $\theta^{(i)}_{\{0,l\}}$, that is, the parameter vector obtained by modifying $\theta^{(i)}_{\{0\}}$ to allow the standard deviation of the l^{th} shock shock to be equal to the corresponding value of $\theta^{(i)}$. We then compute

$$v^{(i)}\big(j, h, \{0, l\}\big) = V\big(y_{it+h}|y_{1:t}, \theta^{(i)}_{\{0,l\}}\big)$$

  and we isolate the effect of shock l by subtracting the effect of the initial condition as follows:

$$v^{(i)}\big(j, h, \{l\}\big) = v^{(i)}\big(j, h, \{0, l\}\big) - v^{(i)}\big(j, h, \{0\}\big)$$

  and the corresponding FEVD coefficients are

$$\text{FEVD}^{(i)}\big(j, h, \{l\}\big) = 100 \times \frac{v^{(i)}\big(j, h, \{l\}\big)}{v^{(i)}\big(j, h, \{all\}\big)}.$$

- The FEVD coefficients describe above by construction sum to 1 across all sources of uncertainty for each variable (initial condition, continuous shocks, measurement errors, and variance switches).

- These computations are repeated for all draws from the posterior distribution and results are averaged across draws. In Table 2 of the paper, we report the posterior means of each FEVD coefficient (one for each variable and for each shock) and the corresponding 5% and 95% quantiles at different horizons.

Appendix C: Nonexpected utility, habit persistence, and expected excess holding period returns

As mentioned in Section 4.4 in the paper, we conjectured that our estimates for strong habit persistence ($h = 0.82$) and in support of nonexpected utility specification ($\gamma = 7.14$, much above the estimate of $\psi = 1.49$), are crucially important to determine the level and volatility of estimated bond excess holding period returns and premia. To confirm our conjecture, we run two sensitivity experiments:

1. we compare filtered excess holding period returns using the mean of the posterior distribution of our parameters (`baseline`) with those obtained by setting $h = 0$ (`no habits`) and keeping all other parameters at their value in the baseline;
2. we compare baseline filtered expected excess holding period returns with those obtained setting $\gamma = \psi$, as in a `power utility` function, and keeping the other parameters at their baseline value.

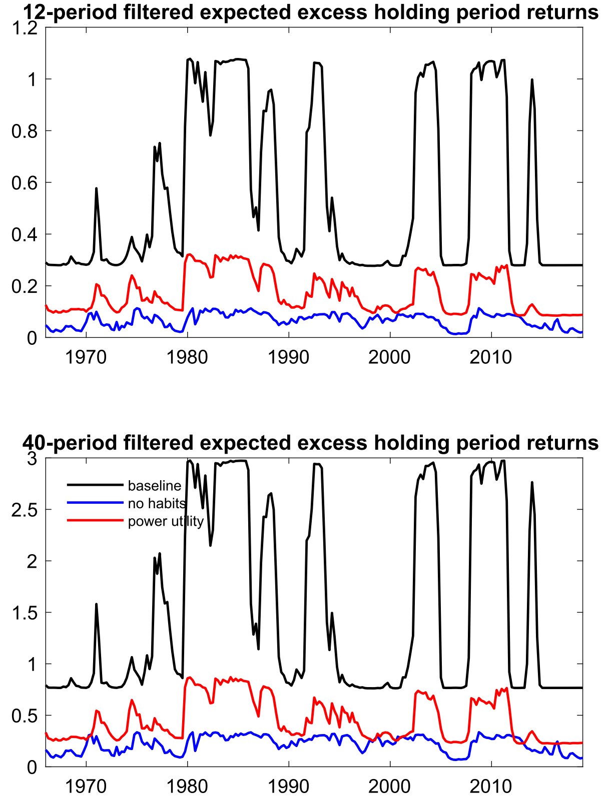

Figure C.1. Filtered expected excess holding period returns: baseline (posterior mean), no habits, power utility.

As shown in both panels of Figure C.1, both the `no habits` and `power utility` parameter configurations produce substantial drops in the level and the volatility in expected excess holding period returns.

References

Amisano, Gianni and Oreste Tristani (2011), "Exact likelihood computation for nonlinear DSGE models with heteroskedastic innovations." *Journal of Economic Dynamics and Control*, 35, 2167–2185. [12]

An, Sungbae and Frank Schorfheide (2007), "Bayesian analysis of DSGE models." *Econometric Reviews*, 26, 113–172. [16]

Bianchi, Francesco (2016), "Methods for measuring expectations and uncertainty in Markov-switching models." *Journal of Econometrics*, 190, 79–99. [18]

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber (2011), "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins." *American Economic Review Papers and Proceedings*, 101, 471–475. [11]

Christiano, Lawrence J., Roberto Motto, and Massimo Rostagno (2014), "Risk shocks." *American Economic Review*, 104, 27–65. [11]

Goffe, William L., Gary D. Ferrier, and John. Rogers (1994), "Global optimization of statistical functions with simulated annealing." *Journal of Econometrics*, 190, 79–99. [16]

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007), "The U.S. Treasury yield curve: 1961 to the present." *Journal of Monetary Economics*, 54, 2291–2304. [11]

Kim, Chang-Jin (1994), "Dynamic linear models with Markov-switching." *Journal of Econometrics*, 60, 1–22. [13]

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter (1996), "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, 74, 119–147. [18]

Maddison, Angus (2003), *The world economy: historical statistics*. OECD Publishing, Paris. [11]

Schmitt-Grohé, Stephanie and Martín Uribe (2004), "Optimal fiscal and monetary policy under sticky prices." *Journal of Economic Theory*, 114, 198–230. [11]

Trabandt, Mathias and Harald Uhlig (2011), "The Laffer curve revisited." *Journal of Monetary Economics*, 58, 305–327. [4, 8]

Co-editor Tao Zha handled this manuscript.

Manuscript received 4 February, 2019; final version accepted 27 September, 2022; available online 18 October, 2022.