

# Monetary policy and long- term interest rates

## Technical appendix – not for publication

Gianni Amisano\*  
Board of Governors  
of the Federal Reserve System

Oreste Tristani†  
ECB and CEPR

September 26, 2022

### Abstract

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/gianniamisanowebiste/>

## A Model details

### A.1 The household problem

The optimization problem is:

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

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

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

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

$$J(W_t) = \max \left\{ (1 - \beta) u_t^{1-\psi} + \beta \left[ E_t J^{1-\gamma}(W_{t+1}) \right]^{\frac{1-\psi}{1-\gamma}} \right\}^{\frac{1}{1-\psi}} + \\ - \Lambda_t \left[ P_t C_t + E_t Q_{t,t+1} W_{t+1} - W_t - w_t N_t - \int_0^1 \Psi_t(i) di + T_t \right],$$

---

\*Email: [gianni.amisano@frb.gov](mailto:gianni.amisano@frb.gov).

†Email: [oreste.tristani@ecb.europa.eu](mailto:oreste.tristani@ecb.europa.eu).

where

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

and

$$\frac{\partial N_t(j)}{\partial w_t(j)} = -\theta_{w,t} \frac{N_t(j)}{w_t(j)}.$$

Using the aggregator function

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

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

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

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

$$\frac{\Lambda_t P_t}{u_{c,t}} = U_{u,t},$$

$$\frac{u_{N,t}}{u_{c,t}} = \frac{1 - \theta_{w,t} \frac{w_t(j)}{P_t}}{\theta_{w,t}},$$

$$Q_{t,t+1} = U_{v,t} [\mathbf{E}_t J_{t+1}^{1-\gamma}]^{\frac{\gamma}{1-\gamma}} J_{t+1}^{-\gamma} \frac{\Lambda_{t+1}}{\Lambda_t},$$

or

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

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

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

## A.2 Detrending

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

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

## A.3 Expected excess holding period returns

Define zero-coupon bond prices as

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

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

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

or to second order

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

Expected holding period returns on a  $n$ -period bonds are

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

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

### 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  $\hat{h}_{n,t}$ , we obtain

$$\hat{h}_{n,t} + \frac{1}{2} \hat{h}_{n,t}^2 = \mathbf{E}_t \hat{b}_{n-1,t+1} - \hat{b}_{n,t} + \frac{1}{2} \mathbf{E}_t \hat{b}_{n-1,t+1}^2 + \frac{1}{2} \hat{b}_{n,t}^2 - \hat{b}_{n,t} \mathbf{E}_t \hat{b}_{n-1,t+1}$$

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

$$\hat{h}_{n,t} = -\hat{b}_{n,t} + \mathbf{E}_t \hat{b}_{n-1,t+1} + \frac{1}{2} \text{Var}_t \hat{b}_{n-1,t+1},$$

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

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

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

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

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

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

and approximated as

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

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

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

We therefore obtain

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

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

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

and to first order

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

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

Using this expression in the excess holding period return we obtain

$$\widehat{h}_{n,t} - \widehat{i}_t = \text{Cov}_t \left[ \widehat{b}_{t+1,n-1}, -\frac{\psi}{1-h} \widehat{c}_{t+1} - \psi \left( 1 + \frac{1}{\phi} \right) \frac{\bar{n}}{1-\bar{n}} \widehat{l}_{t+1} - \psi \widehat{\xi}_{t+1} - \widehat{\pi}_{t+1} - (\gamma-\psi) (\widehat{\xi}_{t+1} + \widehat{j}_{t+1}) \right].$$

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  $\widehat{\xi}_{t+1} + \widehat{j}_{t+1}$ )

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

It follows that

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

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

$$\widehat{h}_{n,t} - \widehat{i}_t = \widetilde{\sigma}^2 F_{B_{n-1}} \text{E}_t [u_{t+1} u'_{t+1}] \left( \psi \frac{1}{1-h} F_c + \psi \left( 1 + \frac{1}{\phi} \right) \frac{\bar{n}}{1-\bar{n}} F_l + \psi F_\xi + F_\pi + (\gamma - \psi) F_j \right)'$$

#### 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

$$B_{t,n}^{EH} = \frac{1}{I_t} \text{E}_t [B_{t+1,n-1}^{EH}]$$

with 1-period yields defined as

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

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

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

#### A.5 Firms' optimization problem

Under Rotemberg prices, firm  $j$  maximizes real profits

$$\max_{P_t^j} \text{E}_t \sum_{s=t}^{\infty} Q_{t,s} \left[ \frac{P_s^j Y_s^j}{P_s} - \frac{w_s}{P_s} \left( \frac{Y_s^j}{A_s} \right)^{\frac{1}{\alpha}} - \frac{\zeta}{2} \left( \frac{P_s^j}{P_{s-1}^j} - \left( \Pi_{s\pi,s}^* \right)^{1-\iota} \Pi_{s-1}^\iota \right)^2 Y_s \right],$$

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 labour index defined above.

The FOC is

$$0 = (1 - \theta) \left( \frac{P_t^j}{P_t} \right)^{-\theta} Y_t \frac{1}{P_t} + \frac{\theta}{\alpha} \frac{w_t}{P_t} \left( \frac{Y_t}{A_t} \right)^{\frac{1}{\alpha}} \left( \frac{P_t^j}{P_t} \right)^{-\frac{\theta}{\alpha} - 1} \frac{1}{P_t} - \zeta \left( \frac{P_t^j}{P_{t-1}^j} - \left( \Pi_{s_{\pi,t}}^* \right)^{1-\iota} \Pi_{t-1}^\iota \right) Y_t \frac{1}{P_{t-1}^j} +$$

$$+ E_t Q_{t,t+1} \zeta \left( \frac{P_{t+1}^j}{P_t^j} - \left( \Pi_{s_{\pi,t+1}}^* \right)^{1-\iota} \Pi_t^\iota \right) Y_{t+1} \frac{P_{t+1}^j}{P_t^j} \frac{1}{P_t^j},$$

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

$$(\theta - 1) \tilde{Y}_t + \zeta \left( \Pi_t - \left( \Pi_{s_{\pi,t}}^* \right)^{1-\iota} \Pi_{t-1}^\iota \right) \tilde{Y}_t \Pi_t = \frac{\theta}{\alpha} \frac{\tilde{w}_t}{P_t} \frac{1}{Z_t^{\frac{1}{\alpha}}} \tilde{Y}_t^{\frac{1}{\alpha}}$$

$$+ E_t Q_{t,t+1} \zeta \left( \Pi_{t+1} - \left( \Pi_{s_{\pi,t+1}}^* \right)^{1-\iota} \Pi_t^\iota \right) \tilde{Y}_{t+1} \Xi_{t+1} \Pi_{t+1}.$$

## A.6 Equilibrium

Equilibrium is described by the following system:

- households

$$\frac{\Lambda_t P_t}{u_{c,t}} = (1 - \beta) \tilde{u}_t^{-\psi} \tilde{J}_t^\psi,$$

$$\frac{\tilde{u}_{N,t}}{u_{c,t}} = \frac{\theta_{w,t} - 1}{\theta_{w,t}} \frac{\tilde{w}_t}{P_t},$$

$$\tilde{J}_t^{1-\psi} = (1 - \beta) \tilde{u}_t^{1-\psi} + \beta \left[ E_t \Xi_{t+1}^{1-\gamma} \tilde{J}_{t+1}^{1-\gamma} \right]^{\frac{1-\psi}{1-\gamma}},$$

$$\tilde{u}_t = u \left( \tilde{C}_t - h \tilde{C}_{t-1}, 1 - N_t \right),$$

$$Q_{t,t+1} = \beta \left[ E_t \tilde{J}_{t+1}^{1-\gamma} \Xi_{t+1}^{1-\gamma} \right]^{\frac{\gamma-\psi}{1-\gamma}} \frac{\tilde{J}_t^\psi}{\tilde{J}_{t+1}^\gamma \Xi_{t+1}^\gamma} \frac{\Lambda_{t+1}}{\Lambda_t};$$

- firms

$$(\theta - 1) \tilde{Y}_t = -\zeta \left( \Pi_t - \left( \Pi_{s_{\pi,t}}^* \right)^{1-\iota} \Pi_{t-1}^\iota \right) \tilde{Y}_t \Pi_t + \frac{\theta}{\alpha} \frac{\tilde{w}_t}{P_t} \frac{1}{Z_t^{\frac{1}{\alpha}}} \tilde{Y}_t^{\frac{1}{\alpha}} +$$

$$+ E_t Q_{t,t+1} \zeta \left( \Pi_{t+1} - \left( \Pi_{s_{\pi,t+1}}^* \right)^{1-\iota} \Pi_t^\iota \right) \tilde{Y}_{t+1} \Xi_{t+1} \Pi_{t+1};$$

- market clearing

$$\tilde{Y}_t = \tilde{C}_t + \tilde{G}_t + \frac{\zeta}{2} \left( \Pi_t - \left( \Pi_{s_{\pi,t}}^* \right)^{1-\iota} \Pi_{t-1}^\iota \right)^2 \tilde{Y}_t,$$

$$N_t = \tilde{Y}_t^{\frac{1}{\alpha}} Z_t^{-\frac{1}{\alpha}};$$

- policy rule

$$I_t = \left( \frac{\Pi_{s_{\pi,t}}^* \Xi_t^\psi}{\beta} \right)^{1-\rho_I} \left( \frac{\Pi_t}{\Pi_{s_{\pi,t}}^*} \right)^{\psi_\Pi} \left( \frac{\tilde{Y}_t}{\tilde{Y}} \right)^{\psi_Y} I_{t-1}^{\rho_I} e^{\eta_{t+1}},$$

$$\ln \Pi_{s_{\pi,t}}^* = s_{\pi,t} \ln \Pi_{s_{\pi,H}}^* + (1 - s_{\pi,t}) \ln \Pi_{s_{\pi,L}}^*;$$

- shocks

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

- standard deviations

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

- $C_{-1}, I_{-1}, \Pi_{-1}$  given.

## A.7 Numerical implementation

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

$$\begin{aligned}\tilde{D}_t &= \frac{\text{E}_t \Xi_{t+1}^{1-\gamma} \tilde{J}_{t+1}^{1-\gamma}}{\kappa^{1-\gamma}}, \\ \tilde{J}_t^{1-\psi} &= (1-\beta) \tilde{u}_t^{1-\psi} + \beta \kappa^{1-\psi} \tilde{D}_t^{\frac{1-\psi}{1-\gamma}}, \\ Q_{t,t+1} &= \beta \left( \frac{\kappa \tilde{D}_t^{\frac{1}{1-\gamma}}}{\tilde{J}_{t+1}} \right)^{\gamma-\psi} \left( \frac{\tilde{u}_{t+1}}{\tilde{u}_t} \right)^{-\psi} \frac{u_{c,t+1}}{u_{c,t}} \frac{1}{\Xi_{t+1}^\gamma} \frac{1}{\Pi_{t+1}}.\end{aligned}$$

## A.8 Functional forms

We rely on the Trabandt and Uhlig (2011) form for temporary utility, i.e.

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

As a result

$$\begin{aligned}\frac{\tilde{w}_t}{P_t} &= \frac{\eta\psi \left( 1 + \frac{1}{\phi} \right) (\tilde{C}_t - h\tilde{C}_{t-1}) N_t^{\frac{1}{\phi}}}{1 - \eta(1-\psi) N_t^{1+\frac{1}{\phi}}} \frac{\theta_{w,t}}{\theta_{w,t} - 1}, \\ \tilde{J}_t^{1-\psi} &= (1-\beta) (\tilde{C}_t - h\tilde{C}_{t-1})^{1-\psi} \left( 1 - \eta(1-\psi) N_t^{1+\frac{1}{\phi}} \right)^\psi + \beta \kappa^{1-\psi} \tilde{D}_t^{\frac{1-\psi}{1-\gamma}}, \\ Q_{t,t+1} &= \beta \left( \frac{\kappa \tilde{D}_t^{\frac{1}{1-\gamma}}}{\tilde{J}_{t+1}} \right)^{\gamma-\psi} \left( \frac{\tilde{C}_{t+1} - h\tilde{C}_t}{\tilde{C}_t - h\tilde{C}_{t-1}} \right)^{-\psi} \left( \frac{1 - \eta(1-\psi) N_{t+1}^{1+\frac{1}{\phi}}}{1 - \eta(1-\psi) N_t^{1+\frac{1}{\phi}}} \right)^\psi \frac{1}{\Xi_{t+1}^\gamma} \frac{1}{\Pi_{t+1}},\end{aligned}$$

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

## 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 labour supply constant.

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

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

can be written as<sup>1</sup>

$$\hat{q}_{t,t+1} = -\psi \Delta \widehat{\overleftarrow{c}}_{t+1} - \psi \left( 1 + \frac{1}{\phi} \right) \frac{\bar{n}}{1 - \bar{n}} \Delta \hat{N}_{t+1} - \psi \hat{\xi}_{t+1} - \hat{\pi}_{t+1} - (\gamma - \psi) \left( \hat{\xi}_{t+1} + \hat{j}_{t+1} - E_t \left[ \hat{\xi}_{t+1} + \hat{j}_{t+1} \right] \right),$$

where

$$\hat{j}_t + \hat{\xi}_t = \sum_{i=0}^{\infty} \left( \beta \Xi^{1-\psi} \right)^i E_t \left[ \hat{\xi}_{t+i} + \left( 1 - \beta \Xi^{1-\psi} \right) \left( \widehat{\overleftarrow{c}}_{t+i} - \frac{\psi}{1-\psi} \left( 1 + \frac{1}{\phi} \right) \frac{\bar{n}}{1-\bar{n}} \hat{N}_{t+i} \right) \right].$$

As a result,

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

and the real rate is

$$\hat{r}_t = \psi E_t \Delta \widehat{\overleftarrow{c}}_{t+1} + \psi \left( 1 + \frac{1}{\phi} \right) \frac{\bar{n}}{1 - \bar{n}} E_t \Delta \hat{N}_{t+1} + \psi E_t \hat{\xi}_{t+1}.$$

Rearranging terms

$$\widehat{\overleftarrow{c}}_t = -\frac{1}{\psi} \hat{r}_t + E_t \widehat{\overleftarrow{c}}_{t+1} + \frac{1}{\psi} \left( 1 + \frac{1}{\phi} \right) \frac{\bar{n}}{1 - \bar{n}} E_t \Delta \hat{N}_{t+1} + E_t \hat{\xi}_{t+1},$$

so that the long-run elasticity of substitution  $\overline{EIS}$ , i.e. 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

$$\widehat{c}_t = -\frac{1}{\psi} \frac{1-h}{1+h} \hat{r}_t + \frac{1}{1+h} E_t \widehat{c}_{t+1} + \frac{h}{1+h} \widehat{c}_{t-1} + \frac{1-h}{1+h} \left( 1 + \frac{1}{\phi} \right) \frac{\bar{n}}{1-\bar{n}} E_t \Delta \hat{N}_{t+1} + \frac{1-h}{1+h} E_t \hat{\xi}_{t+1}.$$

<sup>1</sup>In these derivations,  $\kappa = 1$ .



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

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

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

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

i.e. it includes two components. The first component, as before, is a complete portfolio of state-contingent bonds  $W_t(j)$  yielding a return  $E_t Q_{t,t+1} W_{t+1}(j)$  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  $i$

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

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

$$\begin{aligned} P_t^S &= \int_0^1 P_t^S(i) di \\ \delta_t &= \int_0^1 \delta_t(i) di \end{aligned}$$

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

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

Equilibrium dividends can be obtained from households' total income

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

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

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

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

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

Note that in the non-stochastic steady state

$$ERP^n = 0$$

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

$$\bar{R}_t = \frac{I_t}{E_t \Pi_{t+1}}$$

## B Model estimation

### B.1 Data definition and sources

The macroeconomic data series we use for estimation are obtained from FRED <sup>2</sup> and their mnemonics and original sources are as follows:

- Civilian Non-inst. Population: CNP160V;
- 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).<sup>3</sup>

### B.2 Prior specification details

For the  $\phi$  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 labour supply (see e.g. the evidence reviewed in Chetty *et al.*, 2011). We use a shifted Gamma distribution for  $\psi$  and  $\gamma$ , to ensure that  $\psi, \gamma > 1$ . We centre the distribution of  $\psi$  around a value above but close to 1. For the  $\gamma$  parameter, which contributes to shape risk aversion, we use a very large standard deviation whose prior 95 percent confidence set goes from 2 to 30. The habit parameter has a beta prior centred around 0.5. Finally, for  $\beta$  we use a relatively tight prior with a mean of 0.9985. This is consistent with assumptions made in models with growth – see e.g. Christiano, Motto and Rostagno (2014).

<sup>2</sup><https://fred.stlouisfed.org/>.

<sup>3</sup><https://www.federalreserve.gov/data/yield-curve-tables/feds200628.csv>.

For the long run parameter  $\Xi$  we rely on a more dogmatic prior. We use a tight prior centred around an annualized value of 2 percent, which is consistent with the average per-capita U.S. GDP/GNP growth from the 1870s to the 1950s – see Maddison (2013).

The price adjustment cost  $\zeta$  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 centre the prior around 15, which is roughly consistent, for example, with the value used in Schmitt-Grohé and Uribe (2004b), but allow for a relatively large standard deviation. For inflation indexation, we rely on a beta prior centred around 0.5.

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

### B.3 Likelihood computation

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

$$y_{t+1}^o = k_{y,j} + F\hat{x}_{t+1} + \frac{1}{2} \left( I_{n_y} \otimes \hat{x}'_{t+1} \right) E\hat{x}_{t+1} + Dv_{t+1}, \quad (1)$$

$$\hat{x}_{t+1} = k_{x,i} + P\hat{x}_t + \frac{1}{2} \left( I_{n_x} \otimes \hat{x}'_t \right) G\hat{x}_t + \tilde{\sigma}\Sigma_i w_{t+1}, \quad (2)$$

$$s_t \sim MS(Q), \quad (3)$$

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  $y_t^o$  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  $\Sigma_i$  (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

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

where

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

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

$$\begin{aligned}
\hat{x}_{t+1|t}^{(i,j)} &= \tilde{k}_{x,t}^{(i)} + \tilde{P}_t^{(i)} \hat{x}_{t|t}^{(i)} = \hat{x}_{t+1|t}^{(i)}, \\
Q_{t+1|t}^{(i,j)} &= \tilde{P}_t^{(i)} Q_{t|t}^{(i,j)} \tilde{P}_t^{(i)'} + \Sigma_i \Sigma_i' = Q_{t+1|t}^{(i)},
\end{aligned}$$

and update the vector of continuous state variables

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

$$\begin{aligned}
Q_{t+1|t+1}^{(j)} &= \sum_{i=1}^m \left[ \left( \hat{x}_{t+1|t+1}^{(i,j)} - \hat{x}_{t+1|t+1}^{(j)} \right) \left( \hat{x}_{t+1|t+1}^{(i,j)} - \hat{x}_{t+1|t+1}^{(j)} \right)' + Q_{t+1|t+1}^{(i,j)} \right] \times \\
&\quad \times p(s_t = i | s_{t+1} = j, \underline{y}_{1:t+1}),
\end{aligned}$$

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:

- $x_t^{(i)}, i = 1, 2, \dots, N$ , 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  $x_0^{(i)}, i = 1, 2, \dots, N$ , from its ergodic distribution;
- for each  $t = 1, 2, \dots, T$ ,
  - draw  $x_t^{(i)}, i = 1, 2, \dots, N$  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  $\tilde{p}(y_t|x_{t-1}^{(i)}, \hat{\theta})$ , i.e. 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 step-wise EKF linearization. We call the resulting likelihood "EKF+SMC".

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 (1) reports the correlation among log likelihood computations throughout the sample (that is,  $\log(p(y_t|y_{t-1}, \hat{\theta}))$ ,  $t = 1, 2, \dots, T$ ) across the three methods, and shows very high, although not perfect correlations. In particular, as Figure (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 (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 `csmminwell`) 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 allow us to compute first and second order moments analytically via a pruning approach, i.e. 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 (1), (2) and (3). From the state space representation we obtain the unconditional covariance matrix of state vector shocks as

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

where  $\pi_i$  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  $\hat{x}_t$ , as solution of

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

Dynamic covariance matrices are obtained by applying the recursion

$$Cov(\hat{x}_t, \hat{x}_{t-j}) = \Omega_{xx,j} = P\Omega_{xx,j-1}, j = 1, 2, 3, \dots$$

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

$$\begin{aligned} \Omega_{yy,0} &= Cov(y_t^o) = F\Omega_{xx,0}F' + DD', \\ \Omega_{yy,j} &= Cov(y_t^o, y_{t-j}^o) = F\Omega_{xx,j}F', j = 1, 2, 3, \dots \end{aligned}$$

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

$$y_{t+1}^o = k_{y,j} + F\hat{x}_{t+1} + \frac{1}{2}\bar{E}vec(\hat{x}_{t+1}\hat{x}_{t+1}') + Dv_{t+1}, \quad (4)$$

$$\hat{x}_{t+1} = k_{x,i} + P\hat{x}_t + \frac{1}{2}\bar{G}vec(\hat{x}_t\hat{x}_t') + \tilde{\sigma}\Sigma_i w_{t+1}, \quad (5)$$

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

$$\begin{aligned} \mu_y &= E(y_t^o) = k_y + F\mu_x + \frac{1}{2}\bar{E}vec(\Omega_{xx,0}), k_y = \sum_{j=1}^m k_{y,j}\pi_j, \\ \mu_x &= E(\hat{x}_t) = [I_{n_x} - P]^{-1}[k_x + \frac{1}{2}\bar{G}vec(\Omega_{xx,0})], k_x = \sum_{i=1}^m k_{x,i}\pi_i. \end{aligned}$$

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

### B.5.2 Regime switching affecting the non-stochastic 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 in terms of the continuous states only but its linear slope terms will be regime-specific. In this case, the computations of moments become 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. ergodic mean. The solution can be re-written by with constant This implies that the to the model's solution has

### 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, i.e. the shocks that lead to a change in the discrete Markov switching process that affects the model. We define  $\varepsilon_t$  as the vector containing all the shocks affecting continuous and discrete states. We compute IRFs to a shock  $\varepsilon_{jt}$  of size  $\delta_j$  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  $\hat{x}_t^{(i)}, s_t^{(i)}$  from their joint posterior distribution conditional on  $\theta^{(i)}$ ;
- draw two histories of shocks  $\varepsilon_{t+h}^{(i,1)}, \varepsilon_{t+h}^{(i,2)}, h = 0, 1, 2, \dots, H$ , which are totally identical but differ only for the shock  $\varepsilon_{jt}$ , 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  $\hat{x}_t^{(i)}, s_t^{(i)}$ , and generate 2 paths

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

the difference between these two path traces the dynamic response of shock  $\delta_j$ ;

- 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  $s_{t+h}, h = 0, 1, 2, \dots, H$  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, i.e. 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 2, 3, 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 non linear ones, such as the quadratic MS-DSGE model used in our paper. In particular, difficulties arise since:

1. the model is non-linear;
2. there are shocks in the variances, i.e. 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 non-linearities 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, i.e. taking into consideration only linear terms. In order to describe how the variance decomposition results contained in the paper are computed, we define

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

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 modifying from  $\theta^{(i)}$  to  $\theta_{\{0\}}^{(i)}$  setting all shocks standard deviations 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  $\varepsilon_t^k$ ,  $k = 1, 2, 3, 4, 5$ , 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, i.e. 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

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

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

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

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

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

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

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

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

and

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

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

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

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:

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

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

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

- For each of the 8 continuous shocks without Markov-switching variance, i.e. the mark-up shock  $\varepsilon_t^\mu$ , the permanent technology shock  $\varepsilon_t^\xi$  and the 6 measurement errors ( $l = 1, 2, \dots, 8$ ), we measure the FEVD contribution by defining  $\theta_{\{0,l\}}^{(i)}$ , i.e. the

parameter vector obtained by modifying  $\theta_{\{0\}}^{(i)}$  to allow the standard deviation of the  $l^{th}$  shock to be equal to the corresponding value of  $\theta^{(i)}$ . We then compute

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

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

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

and the corresponding FEVD coefficients are

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

- 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.

## C Non-expected 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 non-expected 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.

. As shown in both panels of Figure (), both the **no habits** and **power utility** parameter configurations produce substantial drop in the level and the volatility in expected excess holding period returns.

## References

- [1] An, S. and F. Schorfheide (2007): "Bayesian analysis of DSGE models", *Econometric Reviews*, 26, 223-172.

- [2] Bianchi, F. (2016): "Methods for measuring expectations and uncertainty in Markov-switching models", *Journal of Econometrics*, 26, 223-172.
- [3] 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, 471-75.
- [4] Goffe, W., G. Ferrier and J. Rogers (1994): "Global optimization of statistical functions with simulated annealing", *Journal of Econometrics*, 190, 79-99.
- [5] Koop, G., M. Pesaran and S. Potter (1996): "Impulse response analysis in nonlinear multivariate models", *Journal of Econometrics*, 74, 119-147.

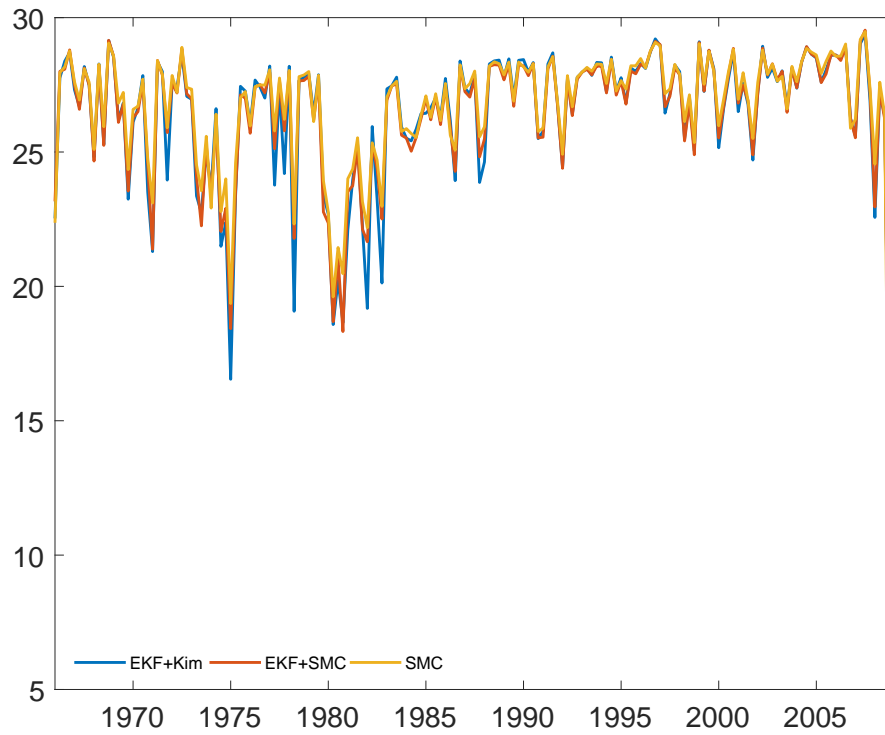


Figure 1: Likelihood computations: EKF+Kim, EKF+SMC, SMC

	Kim+EKF	EKF+SMC	SMC
EKF+Kim	1.00	0.98	0.96
EKF+SMC	0.98	1.00	0.98
SMC	0.96	0.98	1.00

Table 1: Likelihood computations comparisons: Kim+EKF, EKF+SMC, full SMC

Comparison filtered latent states, EKF+Kim v. SMC

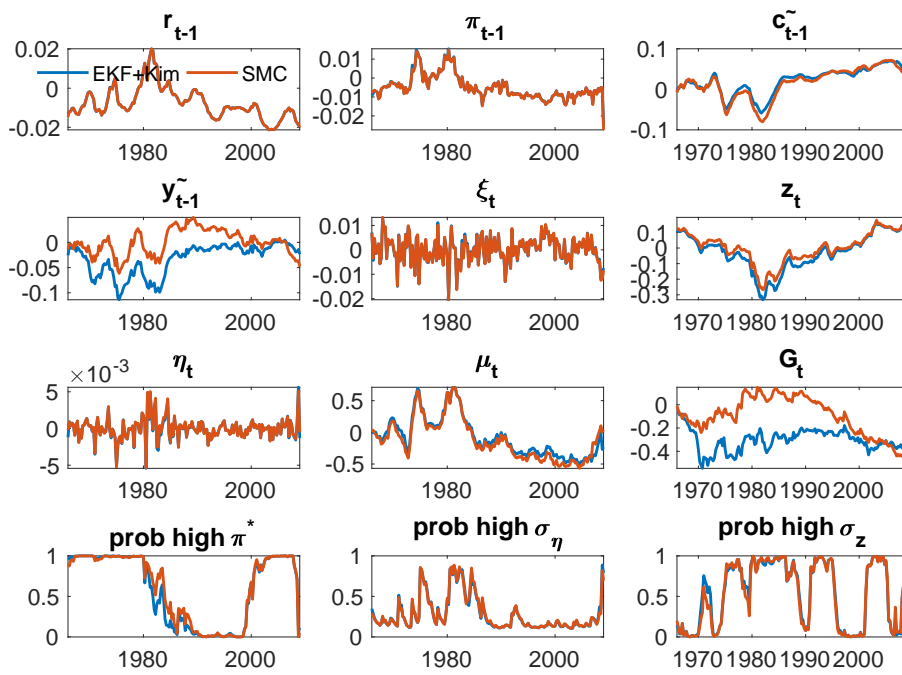


Figure 2: Latent variables computations: EKF+Kim v. SMC

	$\xi$	$z$	$pol$	$\mu$	$G$	$s_{\pi^*}$	$s_{pol}$	$s_z$	$meas.err.$
$\pi_t$	5.8	12.5	6.3	36.0	9.8	22.1	4.4	3.2	0.0
$\Delta c$	58.3	0.0	0.9	1.0	0.3	0.3	0.6	0.6	38.0
$\Delta y$	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
$ERP_{n_t}$	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 2: Variance decomposition 1 steps ahead

	$\xi$	$z$	$pol$	$\mu$	$G$	$s_{\pi^*}$	$s_{pol}$	$s_z$	$meas.err.$
$\pi_t$	5.5	9.0	6.2	26.3	8.4	18.0	12.6	14.0	0.0
$\Delta c$	53.0	1.7	0.5	6.6	0.5	1.2	1.6	1.8	33.2
$\Delta y$	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
$ERP_{n_t}$	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

Table 3: Variance decomposition 12 steps ahead

Legend:  $\pi$ ,  $\Delta y$ ,  $i$ ,  $Rbar$ ,  $R_n$ ,  $EHR40$ ,  $R40$ ,  $ERP$ ,  $PiEqbar\_real$  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.

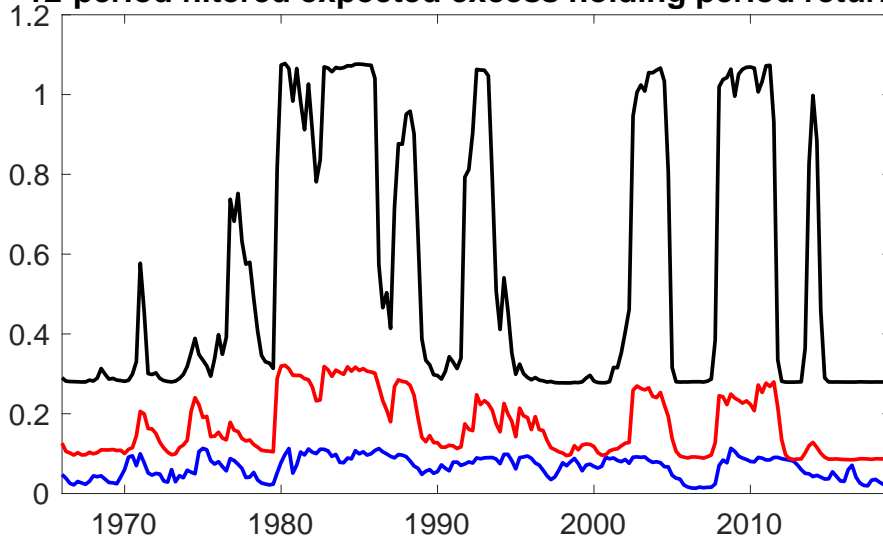
	$\xi$	$z$	$pol$	$\mu$	$G$	$s_{\pi^*}$	$s_{pol}$	$s_z$	$meas.err.$
$\pi_t$	4.9	9.2	6.0	23.0	7.4	17.1	15.1	17.2	0.0
$\Delta c$	50.0	0.6	1.2	7.4	0.6	1.4	2.6	2.4	33.7
$\Delta y$	35.3	1.4	0.8	6.6	22.0	0.9	2.6	2.0	28.3
$r$	3.0	17.4	2.5	10.3	11.6	13.7	7.1	34.5	0.0
$Rbar_t$	11.4	3.7	12.1	16.9	6.0	5.2	32.4	12.4	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
$ERP_{n_t}$	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

Table 4: Variance decomposition 40 steps ahead

Legend:  $\pi$ ,  $\Delta y$ ,  $i$ ,  $Rbar$ ,  $R_n$ ,  $EHR40$ ,  $R40$ ,  $ERP$ ,  $PiEqbar_{real}$  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.



**12-period filtered expected excess holding period returns**



**40-period filtered expected excess holding period returns**

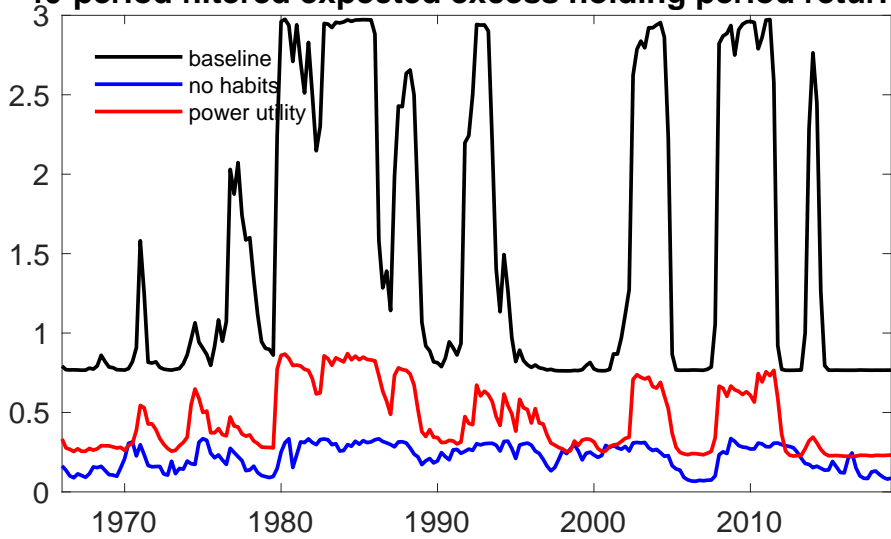


Figure 3: Filtered expected excess holding period returns: baseline (posterior mean), no habits, power utility