Supplement to "Inflation and professional forecast dynamics: An evaluation of stickiness, persistence, and volatility"

(*Quantitative Economics*, Vol. 11, No. 4, November 2020, 1485–1520)

Elmar Mertens
Research Centre, Deutsche Bundesbank

James M. Nason
Department of Economics, NC State University and Centre for Applied Macroeconomic Analysis

This Supplement describes the construction of the state space models estimated in our paper, and a framework for identification. In addition, we provide details about our particle learning filter and particle smoothing algorithms.

S1. SSMs implied by different model variants

This section builds the SSMs that are estimated in Section 4 of the paper. Table 1 of the paper gives an overview of the different SSMs and introduces $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ for each SSM. The SSMs differ by restrictions placed on time-variation in the AR1 inflation gap persistence parameter and sticky-information (SI) weight.

Our baseline SSM is denoted $\mathcal{M}_0$. It has a constant AR1 inflation gap persistence parameter, θ, while the SI weight, λ_t, is time-varying. By letting $\lambda_t = \lambda$, we turn $\mathcal{M}_0$ into $\mathcal{M}_1$ that has λ and θ. However, constructing a recursive state space is more difficult when θ becomes time-varying, θ_t, as in the SSMs $\mathcal{M}_2$ and $\mathcal{M}_3$.

S1.1 *A Stock and Watson UC model of inflation*

We reproduce our version of the SW-UC model of inflation here

$$\pi_t = \tau_t + \tilde{\pi}_t, \tag{A.1}$$

$$\tilde{\pi}_t = \varepsilon_t + \sigma_{\zeta,\pi}\zeta_{\pi,t}, \quad \zeta_{\pi,t} \sim \mathcal{N}(0,1), \tag{A.2}$$

$$\tau_{t+1} = \tau_t + \varsigma_{\eta,t+1}\eta_t, \quad \eta_t \sim \mathcal{N}(0,1), \tag{A.3}$$

$$\varepsilon_{t+1} = \theta\varepsilon_t + \varsigma_{\upsilon,t+1}\upsilon_t, \quad \upsilon_t \sim \mathcal{N}(0,1), |\theta| < 1, \tag{A.4}$$

$$\ln \varsigma^2_{\ell,t+1} = \ln \varsigma^2_{\ell,t} + \sigma_\ell \xi_{\ell,t+1}, \quad \xi_{\ell,t+1} \sim \mathcal{N}(0,1), \ell = \eta, \upsilon, \tag{A.5}$$

Elmar Mertens: elmar.mertens@bundesbank.de
James M. Nason: jmnason@ncsu.edu
The views herein are those of the authors and do not represent the views of the Deutsche Bundesbank or the Eurosystem. Jim Nason thanks the Jenkins Family Economics Fund at North Carolina State University for financial support.

 https://doi.org/10.3982/QE980

where inflation is the sum of trend, τ_t and gap $\tilde{\pi}_t$, which again is the sum of a persistent component, ε_t, and a serially uncorrelated irregular component, $\zeta_{\pi,t}$. Furthermore, $\varsigma_{\eta,t}$, and $\varsigma_{\upsilon,t}$ denote stochastic volatility in the innovation η_t of τ_t, and stochastic volatility in the innovation υ_t of ε_t.

Consider a SW-UC model lacking persistence in gap inflation, $\theta = 0$ (as well as $\zeta_{\pi,t} = 0$), and constant volatility, $\sigma_\eta = \varsigma_{\eta,t}$ and $\sigma_\upsilon = \varsigma_{\upsilon,t}$. The result is a fixed coefficient SW-UC model with an invertible IMA(1, 1) reduced form, $(1 - \mathbf{L})\pi_t = (1 - \varpi\mathbf{L})e_t$, where $\mathbf{L}$ is the lag operator, $\pi_{t-1} = \mathbf{L}\pi_t$, e_t is the one-step ahead forecast error $e_t \equiv \pi_t - \mathbf{E}\{\pi_{t+1}|\pi^t\}$, and the MA1 coefficient satisfies $\varpi \in (0, 1)$.[1] The IMA(1, 1) yields a rational expectations (RE) inflation forecast updating equation, $\mathbf{E}\{\pi_{t+1}|\pi^t\} = (1 - \varpi)\pi_t + \varpi\mathbf{E}\{\pi_t|\pi^{t-1}\}$, where $\pi^t = [\pi_t, \ldots, \pi_1]$ and $K \equiv 1 - \varpi$ is the Kalman gain of trend inflation, $\tau_{t|t} = \tau_{t-1|t-1} + Ke_t$.

Stock and Watson (2007), Grassi and Proietti (2010), and Shephard (2015) noted stochastic volatility in the SW-UC model gives the MA1 coefficient a local time-varying parameter interpretation, ϖ_t. By iterating the RE updating equation $\mathbf{E}\{\pi_{t+1}|\pi^t, \varsigma^t_\eta, \varsigma^t_\upsilon\} = (1 - \varpi_t)\pi_t + \varpi_t\mathbf{E}\{\pi_t|\pi^{t-1}, \varsigma^{t-1}_\eta, \varsigma^{t-1}_\upsilon\}$ backwards, we have an exponentially weighted MA updating recursion or smoother, $\mathbf{E}\{\pi_{t+1}|\pi^t, \varsigma^t_\eta, \varsigma^t_\upsilon\} = \sum_{j=0}^{\infty} \mu_{t,t-j}\pi_{t-j}$, that is a forecasting tool traced to Muth (1960), where the discount $\mu_{t,t} = 1 - \varpi_t$ at $j = 0$ or $\mu_{t,t-j} = \mu_{t,t}\prod_{\ell=0}^{j-1}\varpi_{t-\ell}$ for $j \geq 1$. The exponentially weighted MA smoother yields a term structure of RE inflation forecasts in which $\mu_{t,t-j}$ adjusts to changes in π^t, ς^t_η, and ς^t_υ. When $\theta \neq 0$, the RE term structure becomes an input into computing SI inflation forecasts.

S1.2 *The sticky-information prediction mechanism of inflation*

This section revisits the SI prediction mechanism of inflation, which consists of

$$\pi^{\text{SPF}}_{t,t+h} = \mathbf{F}_t\pi_{t+h} + \sigma_{\zeta,h}\zeta_{h,t}, \quad \zeta_{h,t} \sim \mathcal{N}(0, 1), \tag{A.6}$$

$$\mathbf{F}_t\pi_{t+h} = \lambda_t\mathbf{F}_{t-1}\pi_{t+h} + (1 - \lambda_t)\mathbf{E}_t\pi_{t+h}, \quad h = 1, \ldots, \mathcal{H}, \tag{A.7}$$

$$\lambda_{t+1} = \lambda_t + \sigma_\kappa\kappa_t, \quad \kappa_t \sim \mathcal{N}(0, 1) \text{ s.t. } \lambda_{t+1} \in (0, 1), \tag{A.8}$$

where h belongs to the set of positive integers, $h \in \mathbb{Z}^+$, and λ_{t+1} follows a bounded random walk with shocks drawn from a truncated normal that guarantees $\lambda_{t+1} \in (0, 1)$.[2]

The SI law of motion (A.7) implies a exponentially weighted MA smoother. Repeated lagging and substitution of (A.7) produces the SI-exponentially weighted MA smoother

$$\mathbf{F}_t\pi_{t+h} = \sum_{j=0}^{\infty} \Lambda_{t,t-j}\mathbf{E}_{t-j}\pi_{t+h}, \tag{A.9}$$

---

[1] Stock and Watson (2007), Grassi and Proietti (2010), and Shephard (2015) computed ϖ by equating the first two autocovariances of the IMA(1, 1) and the fixed coefficient SW-UC model: $(1 + \varpi^2)\sigma^2_e = \sigma^2_\eta + 2\sigma^2_\varepsilon$ and $-\varpi\sigma^2_e = -\sigma^2_\varepsilon$. For $\sigma_\eta, \sigma_\varepsilon > 0$, and $q = \sigma^2_\eta/\sigma^2_\varepsilon$, the MA coefficient of the invertible IMA(1, 1) representation is given by $\varpi = 1 + 0.5(q - \sqrt{q^2 + 4q})$.

[2] The innovations to λ_{t+1} are drawn from $\kappa_t \sim \mathcal{TN}(0, 1; -\lambda_t/\sigma_\kappa, (1 - \lambda_t)/\sigma_\kappa)$. $\mathcal{TN}(\mu, \sigma^2; \underline{x}, \overline{x})$ denotes a truncated normal with support between $\underline{x}$ and $\overline{x}$, and mean and variance parameters μ and σ^2.

where the time-varying discount $\Lambda_{t,t-j}$ is $1-\lambda_t$ for $j=0$ and $\Lambda_{t,t-j} = \Lambda_{t,t}\prod_{\ell=0}^{j-1}\lambda_{t-\ell}$ otherwise. The SI-exponentially weighted MA smoother (A.9) nests the RE inflation forecast, $\lim_{\lambda_t\longrightarrow 0}\mathbf{F}_t\pi_{t+h} = \mathbf{E}_t\pi_{t+h}$, and the pure SI inflation forecast, $\lim_{\lambda_t\longrightarrow 1}\mathbf{F}_t\pi_{t+h} = \sum_{j=1}^{\infty}\Lambda_{t,t-j}\mathbf{E}_{t-j}\pi_{t+h} = \mathbf{F}_{t-1}\pi_{t+h}$. The former limit equates $\mathbf{F}_t\pi_{t+h}$ to $\mathbf{E}_t\pi_{t+h}$, as λ_t falls to zero. As λ_t moves to its upper bound, $\mathbf{F}_{t-1}\pi_{t+h}$ becomes the SI inflation forecast because the weight on $\mathbf{E}_t\pi_{t+h}$ decreases while increasing on its lags.

S1.3 *The benchmark SSM, $\mathcal{M}_0$*

This section builds $\mathcal{M}_0$ by combining our SW-UC model of inflation, the SI law of motion of inflation forecasts, and the term structure of average SPF inflation predictions. Our SW-UC model is employed to construct a term structure of RE inflation forecasts. This term structure has a two factor representation driven by τ_t and ε_t. We conjecture and verify state equations for SI trend and gap inflation, $\mathbf{F}_t\tau_t$ and $\mathbf{F}_t\varepsilon_t$, that are consistent with the SI law of motion of inflation forecasts. An implication is the SI states, $\mathbf{F}_t\tau_t$ and $\mathbf{F}_t\varepsilon_t$, are the factors of the term structure of SI inflation forecasts. The SI states eliminate $\mathbf{F}_t\pi_{t+h}$ in the term structure equation (A.6) of $\pi^{\mathrm{SPF}}_{t,t+h}$. As a result, $\mathcal{M}_0$ has seven state variables that we group into $\mathcal{X}_t = [\tau_t\ \varepsilon_t]'$, $\mathbf{F}_t\mathcal{X}_t = [\mathbf{F}_t\tau_t\ \mathbf{F}_t\varepsilon_t]'$, and $\mathcal{V}_t = [\varsigma_{\eta,t}\ \varsigma_{v,t}\ \lambda_t]'$.

As already discussed, constructing $\mathcal{M}_0$ is a multistep process. Start by rewriting the observation equations (A.1) and (A.2) of the SW-UC model

$$\pi_t = \delta_{\mathcal{X}}\mathcal{X}_t + \sigma_{\zeta,\pi}\zeta_{\pi,t}, \tag{A.10}$$

where $\delta_{\mathcal{X}} = [1\ 1]$. Stack the random walk (A.3) of τ_{t+1} on top of equation (A.4), which is the AR(1) of ε_{t+1}, to create the state equations of our SW-UC model

$$\mathcal{X}_{t+1} = \boldsymbol{\Theta}\mathcal{X}_t + \boldsymbol{Y}_{t+1}\mathcal{W}_t, \tag{A.11}$$

where $\boldsymbol{\Theta} = \left[\begin{smallmatrix}1 & 0\\ 0 & \theta\end{smallmatrix}\right]$, $\boldsymbol{Y}_{t+1} = \left[\begin{smallmatrix}\varsigma_{\eta,t+1} & 0\\ 0 & \varsigma_{v,t+1}\end{smallmatrix}\right]$, $\mathcal{W}_t = \left[\begin{smallmatrix}\eta_t\\ v_t\end{smallmatrix}\right]$, and the stochastic volatilities, $\ln\varsigma^2_{\eta,t+1}$ and $\ln\varsigma^2_{v,t+1}$, are random walks as described by equation (A.5).

The term structure of RE inflation forecasts is built using the observation equation (A.10) and state equations (A.11). Iterate the state equations (A.11) h-steps ahead, pass $\mathbf{E}_t\{\cdot\}$ through, and substitute to find $\mathbf{E}_t\mathcal{X}_{t+h} = \boldsymbol{\Theta}^h\mathcal{X}_t$, where $\mathbf{E}_t\{\cdot\}$ conditions on $\boldsymbol{\pi}^t$, $\varsigma_{\eta,t}$, and $\varsigma_{v,t}$. Push the observation equation (A.10) h-steps ahead, apply $\mathbf{E}_t\{\cdot\}$, and substitute for $\mathbf{E}_t\mathcal{X}_{t+h}$, to produce the RE term structure of inflation forecasts $\mathbf{E}_t\pi_{t+h} = \delta_{\mathcal{X}}\boldsymbol{\Theta}^h\mathcal{X}_t$.

Next, the SI law of motion (A.7) suggests the law of motion of SI-states is $\mathbf{F}_t\mathcal{X}_{t+1} = \lambda_t\mathbf{F}_{t-1}\mathcal{X}_{t+1} + (1-\lambda_t)\mathbf{E}_t\mathcal{X}_{t+1}$. The SI-exponentially weighted MA smoother (A.9) is consistent with the law of motion of $\mathbf{F}_t\mathcal{X}_{t+1}$. Iterate the latter law of motion backwards and substitute $\boldsymbol{\Theta}^{h+j}\mathcal{X}_{t-j}$ for $\mathbf{E}_{t-j}\mathcal{X}_{t+h}$ to obtain the exponentially weighted MA smoother of the SI-states, $\mathbf{F}_t\mathcal{X}_{t+h} = \boldsymbol{\Theta}^h\sum_{j=0}^{\infty}\Lambda_{t,t-j}\boldsymbol{\Theta}^j\mathcal{X}_{t-j}$. When $h=0$, the exponentially weighted MA smoother of the SI-states is

$$\mathbf{F}_t\mathcal{X}_t = \sum_{j=0}^{\infty}\Lambda_{t,t-j}\boldsymbol{\Theta}^j\mathcal{X}_{t-j}. \tag{A.12}$$

Pull $\mathcal{X}_t$ out of the infinite sum of (A.12) to find $\mathbf{F}_t\mathcal{X}_t = (1-\lambda_t)\mathcal{X}_t + \sum_{j=1}^{\infty} \Lambda_{t,t-j}\boldsymbol{\Theta}^j\mathcal{X}_{t-j}$. Changing the indexes $j = i+1$ and $\ell = s+1$ converts the infinite sum of the previous expression to $\lambda_t\boldsymbol{\Theta}\sum_{i=0}^{\infty}\Lambda_{t-1,t-i-1}\boldsymbol{\Theta}^i\mathcal{X}_{t-i-1}$. Since this infinite sum equals $\mathbf{F}_{t-1}\mathcal{X}_{t-1}$, substitute $\lambda_t\boldsymbol{\Theta}\mathbf{F}_{t-1}\mathcal{X}_{t-1}$ in the previous expression for $\mathbf{F}_t\mathcal{X}_t$ to see that its law of motion is $\mathbf{F}_t\mathcal{X}_t = (1-\lambda_t)\mathcal{X}_t + \lambda_t\boldsymbol{\Theta}\mathbf{F}_{t-1}\mathcal{X}_{t-1}$. Finally, lead the law of motion of $\mathbf{F}_t\mathcal{X}_t$ forward one period and substitute for $\mathcal{X}_{t+1}$ using the SW-UC model's state equations (A.11) to find

$$\mathbf{F}_{t+1}\mathcal{X}_{t+1} = \lambda_{t+1}\boldsymbol{\Theta}\mathbf{F}_t\mathcal{X}_t + (1-\lambda_{t+1})\boldsymbol{\Theta}\mathcal{X}_t + (1-\lambda_{t+1})\boldsymbol{\Upsilon}_{t+1}\mathcal{W}_t, \tag{A.13}$$

which is the system of SI-state equations. The state equations of $\mathcal{M}_0$ are formed by stacking the state equations (A.11) of $\mathcal{X}_{t+1}$ on top of the SI state equations (A.13)

$$\mathcal{S}_{t+1} = \boldsymbol{\mathcal{A}}_{t+1}\mathcal{S}_t + \boldsymbol{\mathcal{B}}_{t+1}\mathcal{W}_t, \tag{A.14}$$

where $\mathcal{S}_t = \left[\begin{smallmatrix}\mathcal{X}_t \\ \mathbf{F}_t\mathcal{X}_t\end{smallmatrix}\right]$, $\boldsymbol{\mathcal{A}}_{t+1} = \left[\begin{smallmatrix}\boldsymbol{\Theta} & \mathbf{0}_{2\times 2} \\ (1-\lambda_{t+1})\boldsymbol{\Theta} & \lambda_{t+1}\boldsymbol{\Theta}\end{smallmatrix}\right]$, $\boldsymbol{\mathcal{B}}_{t+1} = \left[\begin{smallmatrix}\boldsymbol{\Upsilon}_{t+1} \\ (1-\lambda_{t+1})\boldsymbol{\Upsilon}_{t+1}\end{smallmatrix}\right]$, and $\ln\varsigma^2_{\eta,t+1}$, $\ln\varsigma^2_{\upsilon,t+1}$, and λ_{t+1} evolve as the random walks (A.5) and (A.8). Drift in the SI weight and the stochastic volatilities create nonlinearities in the state equations (A.14). However, $\mathcal{S}_{t+1}$ has linear dynamics conditional on a realization of $\mathcal{V}_{t+1}$.

We construct the observation system of $\mathcal{M}_0$ using the observation equation (A.10) of our SW-UC model, SPF measurement equation (A.6), and RE and SI term structures of inflation forecasts. The former term structure replaces $\mathbf{E}_{t-j}\pi_{t+h}$ with $\delta_{\mathcal{X}}\boldsymbol{\Theta}^h\mathcal{X}_{t-j}$ in the sticky inflation forecast-exponentially weighted MA smoother (A.9) to yield $\mathbf{F}_t\pi_{t+h} = \delta_{\mathcal{X}}\boldsymbol{\Theta}^h\sum_{j=0}^{\infty}\Lambda_{t,t-j}\boldsymbol{\Theta}^j\mathcal{X}_{t-j}$. Up to $\delta_{\mathcal{X}}\boldsymbol{\Theta}^h$, $\mathbf{F}_t\pi_{t+h}$ equals the exponentially weighted MA smoother (A.12) of the SI-states. The result is the term structure of SI inflation forecasts, $\mathbf{F}_t\pi_{t+h} = \delta_{\mathcal{X}}\boldsymbol{\Theta}^h\mathbf{F}_t\mathcal{X}_t$. It eliminates $\mathbf{F}_t\pi_{t+h}$ from the SPF term structure equation (A.6), $\pi^{\mathrm{SPF}}_{t,t+h} = \delta_{\mathcal{X}}\boldsymbol{\Theta}^h\mathbf{F}_t\mathcal{X}_t + \sigma_{\zeta,h}\zeta_{h,t}$, which shows the SI-states are factors of the term structure of average SPF inflation predictions. Put these SPF term structure equations beneath the observation equation (A.10) of our SW-UC model to produce the system of observation equations of $\mathcal{M}_0$

$$\mathcal{Y}_t = \boldsymbol{\mathcal{C}}\mathcal{S}_t + \boldsymbol{\mathcal{D}}\mathcal{U}_t, \tag{A.15}$$

where

$$\mathcal{Y}_t = \begin{bmatrix}\pi_t \\ \pi^{\mathrm{SPF}}_{t,t+1} \\ \vdots \\ \pi^{\mathrm{SPF}}_{t,t+\mathcal{H}}\end{bmatrix}, \qquad \boldsymbol{\mathcal{C}} = \begin{bmatrix}\delta_{\mathcal{X}} & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \delta_{\mathcal{X}}\boldsymbol{\Theta} \\ \vdots & \vdots \\ \mathbf{0}_{1\times 2} & \delta_{\mathcal{X}}\boldsymbol{\Theta}^{\mathcal{H}}\end{bmatrix}, \qquad \boldsymbol{\mathcal{D}} = \begin{bmatrix}\sigma_{\zeta,\pi} & 0 & \dots & 0 \\ 0 & \sigma_{\zeta,1} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \sigma_{\zeta,\mathcal{H}}\end{bmatrix},$$

$$\mathcal{U}_t = [\zeta_{\pi,t} \quad \zeta_{1,t} \quad \dots \quad \zeta_{\mathcal{H},t}]', \quad \text{and} \quad \boldsymbol{\Omega}_{\mathcal{U}} = \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}'.$$

The SSM maps $\mathbf{F}_t\pi_{t+h}$, as it appears in the observation equations (A.15), into a linear combination of SI trend and gap.

S1.4 *SSM when θ_t is time-varying*

Time-variation in θ_t induces a nonlinearity in the AR(1) dynamics of the process for gap inflation, which also affects the process for inflation and the state equation of the

SW-UC model of realized inflation. In this case, the process for gap inflation becomes

$$\varepsilon_{t+1} = \theta_{t+1}\varepsilon_t + \varsigma_{\upsilon,t+1}\upsilon_t, \quad \upsilon_t \sim \mathcal{N}(0,1), \tag{A.16}$$

$$\theta_{t+1} = \theta_t + \sigma_\phi \phi_{t+1}, \quad \phi_{t+1} \sim \mathcal{N}(0,1) \text{ s.t. } \theta_{t+1} \in (-1,1), \tag{A.17}$$

where equation (A.16) replaces equation (1.3) of the paper, equation (A.17) is added to the SW-UC model, and $\mathbf{E}_t$ conditions on date t information that includes θ_t. As with the process for λ_{t+1} in (A.8), shocks θ_{t+1} in (A.17) are drawn from a truncated normal distribution, whose bounds dynamically adjust to ensure $|\theta_{t+1}| < 1$ at all times.

Optimal forecasts of the inflation gap then involve higher moments of the process for θ_t, which breaks the parsimonious structure of the term structure of RE forecasts and the term structure of SI forecasts on which the SSM with constant θ rest. Instead, as common in the literature on RE forecasts generated from time-varying VARs, we approximate inflation gap forecasts by assuming the time-varying coefficient θ_t remains (locally) fixed at its current value (and thus ignoring any future variations in this parameter) when forming expectations; see, for example, Justiniano and Primiceri (2008) and Cogley and Sbordone (2008).[3] This approach can be motivated with the anticipated utility model (AUM) of Kreps (1998); see also Cogley and Sargent (2008).[4] As a result, RE forecasts are approximated by $\mathbf{E}_t\varepsilon_{t+h} \approx \theta_t^h \varepsilon_t$, and $\mathbf{E}_t\pi_{t+h} = \delta_{\mathcal{X}}\mathbf{E}_t\mathcal{X}_{t+h}$ where $\mathbf{E}_t\mathcal{X}_{t+h} \approx \boldsymbol{\Theta}_t^h \mathcal{X}_t$.[5]

Regardless of whether θ_t is constant or not, SI forecasts are a distributed lag polynomial of current and past RE forecasts $\mathbf{F}_t\pi_{t+h} = \sum_{j=0}^{\infty} \Lambda_{t,t-j}\mathbf{E}_{t-j}\pi_{t+h}$ with $\Lambda_{t,t} = 1 - \lambda_t$ and $\Lambda_{t,t-j} = (1-\lambda_{t-j})\prod_{k=0}^{j-1}\lambda_{t-k}$, $\forall j > 0$. As shown in the previous section, when the inflation gap persistence parameter θ is constant, the following relationships hold for SI inflation forecasts and SI nowcasts of trend and gap inflation

$$\mathbf{F}_t\pi_{t+h} = \delta_{\mathcal{X}}\mathbf{F}_t\boldsymbol{\Theta}^h\mathbf{F}_t\mathcal{X}_t, \tag{A.18}$$

$$\mathbf{F}_t\mathcal{X}_t = (1-\lambda_t)\mathcal{X}_t + \lambda_t\boldsymbol{\Theta}\mathbf{F}_{t-1}\mathcal{X}_{t-1}, \tag{A.19}$$

where (A.18) led to the measurement equation (A.15) for SPF forecasts and (A.19) informed the derivation of the transition equation (A.14) for the SI states.

Extending the AUM arguments to the case of θ_t, we approximate the evolution of SI forecasts by evaluating the constant-θ equations (A.18) and (A.19) at date t using $\boldsymbol{\Theta}_t$ in place of $\boldsymbol{\Theta}$

$$\mathbf{F}_t\pi_{t+h} \approx \delta_{\mathcal{X}}\boldsymbol{\Theta}_t^h\mathbf{F}_t\mathcal{X}_t, \tag{A.20}$$

$$\mathbf{F}_t\mathcal{X}_t \approx (1-\lambda_t)\mathcal{X}_t + \lambda_t\boldsymbol{\Theta}_t\mathbf{F}_{t-1}\mathcal{X}_{t-1}. \tag{A.21}$$

[3] Note that since θ_t and υ_t are independent, the exact solution to the optimal forecasting problem involves solving $\mathbf{E}_t\varepsilon_{t+h} = \mathbf{E}_t(\prod_{i=1}^h \theta_{t+i})\varepsilon_t$, $\forall h > 0$. The AUM assumption then amounts to ignoring Jensen-inequality terms and replacing $\mathbf{E}_t(\prod_{i=1}^h \theta_{t+i})$ by $\prod_{i=1}^h(\mathbf{E}_t\theta_{t+i}) = \theta_t^h$.

[4] Cogley and Sargent (2008) contended the AUM assumptions result in decision making that is consistent with Bayesian forecasting. Kreps (1998) argued agents engaging in AUM-like behavior are acting rationally when seeing through to the true model is costly.

[5] Similar to the constant-θ case, we have $\boldsymbol{\Theta}_t = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \theta_t \end{smallmatrix}\right]$.

Thus, for θ_t, we build a state space that approximates the true evolution of RE and SI forecasts with two layers of AUM arguments. First, AUM is applied to form RE forecasts at date t as if θ_t were to remain constant at its current value. Similarly, the system of recursive SI forecasting equations is evaluated as if θ_t is held fixed at its current value for updating SI forecasts in the transition equation (A.21) along with projecting forward the SI forecasts using $\mathbf{F}_t\mathcal{X}_{t+h} = \boldsymbol{\Theta}_t^h \mathbf{F}_t\mathcal{X}_t$.

We obtain a similar SSM as in the constant-θ cases, but replace $\boldsymbol{\Theta}$ with $\boldsymbol{\Theta}_t$

$$\mathcal{S}_{t+1} = \boldsymbol{\mathcal{A}}_{t+1}\mathcal{S}_t + \boldsymbol{\mathcal{B}}_{t+1}\mathcal{W}_t, \tag{A.22}$$

where $\boldsymbol{\mathcal{A}}_{t+1} = \left[\begin{smallmatrix} \boldsymbol{\Theta}_{t+1} & \mathbf{0}_{2\times 2} \\ (1-\lambda_{t+1})\boldsymbol{\Theta}_{t+1} & \lambda_{t+1}\boldsymbol{\Theta}_{t+1} \end{smallmatrix}\right]$ and $\boldsymbol{\mathcal{B}}_{t+1} = \left[\begin{smallmatrix} \boldsymbol{\Upsilon}_{t+1} \\ (1-\lambda_{t+1})\boldsymbol{\Upsilon}_{t+1} \end{smallmatrix}\right]$. As in the constant-$\theta$ case of the paper, the state equations (A.22) show that shocks to λ_t alter only the transition and impulse dynamics of $\mathbf{F}_t\mathcal{X}_t$. In contrast, changes in θ_t shift the transition dynamics of all elements of $\mathcal{S}_t$. The same state space equations apply to $\mathcal{M}_3$, where θ_t is time-varying and λ is constant, by replacing λ_t above with λ.

We complete the SSM by constructing its observation equations. First, replace $\mathbf{F}_t\pi_{t+h}$ in the SPF measurement equation (A.6) with the SI term structure of inflation forecasts (A.20) for $h = 1, \dots, \mathcal{H}$. The observation equation (A.1) for inflation of the SW-UC model is identical to the baseline model of the paper, and we obtain

$$\mathcal{Y}_t = \boldsymbol{\mathcal{C}}_t\mathcal{S}_t + \boldsymbol{\mathcal{D}}\mathcal{U}_t, \quad \text{with } \boldsymbol{\mathcal{C}}_t = \begin{bmatrix} \delta_{\mathcal{X}} & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \delta_{\mathcal{X}}\boldsymbol{\Theta}_t \\ \vdots & \vdots \\ \mathbf{0}_{1\times 2} & \delta_{\mathcal{X}}\boldsymbol{\Theta}_t^{\mathcal{H}} \end{bmatrix},$$

and $\mathcal{Y}_t$, $\boldsymbol{\mathcal{D}}$, $\mathcal{U}_t$, and $\boldsymbol{\Omega}_{\mathcal{U}}$ as defined before.

S2. Identification

This section complements our discussion in Section 2.4 of the paper regarding the identification of parameters and latent states within our SSMs. However, the discussion below gives gives a fuller analysis compared with the intuition about the link between specific observable variables and the estimation of several parameters and states found in the paper. The analysis of this section is limited to the case where time variation in the nonlinear state variables is shut down. Since this imposes linearity on the SSMs, our analysis is grounded in (local) identification of linear SSMs. We especially rely on the framework of Komunjer and Ng (2011). They cast their analysis of identification around UC representations of DSGE models. Their work also applies to UC models in which the mapping from parameters to UC coefficients is not subject to the cross-equation restrictions emanating from a DSGE model. For ease of comparison with their paper, this section represents our SSMs in a notation similar to theirs.

S2.1 *Komunjer–Ng framework*

The identification analysis of Komunjer and Ng (KN) rests on the assumption that measurement and state variables are stationary. We extend their framework to models like ours, where state and measurement dynamics include unit root processes. Throughout, our discussion of the KN framework is limited to their "nonsingular" case, in which there are more shocks than observables. This is true in our model.

Consider the following class of UC models studied by KN:

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}(\Psi)\boldsymbol{X}_t + \boldsymbol{B}(\Psi)\boldsymbol{\varepsilon}_{t+1} \quad \text{and} \quad \boldsymbol{Y}_{t+1} = \boldsymbol{C}(\Psi)\boldsymbol{X}_t + \boldsymbol{D}(\Psi)\boldsymbol{\varepsilon}_{t+1}, \tag{A.23}$$

where $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ are state and measurement vectors, respectively, and $\boldsymbol{A}(\Psi)$, $\boldsymbol{B}(\Psi)$, $\boldsymbol{C}(\Psi)$, $\boldsymbol{D}(\Psi)$, are state space matrices expressed as functions of the underlying parameters Ψ. The pair of equations in (A.23) corresponds to equations (1a) and (1b) in KN. As in KN or the work of Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), this SSM has the "ABCD" form. For ease of notation, we will drop the dependency of state space objects (i.e., $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$) on Ψ unless needed.

The shocks ε_{t+1} are mean zero and serially uncorrelated. Without loss of generality, normalize the variances of the elements of ε_{t+1}, $\mathbf{Var}(\varepsilon_{t+1}) = \boldsymbol{I}$.[6] The coefficient matrices are treated as continuously differentiable in Ψ. We also rely on KN assuming that $\boldsymbol{D}\boldsymbol{D}'$ is positive definite.[7] In addition $(\boldsymbol{A}, \boldsymbol{C})$ are assumed to be detectable under their assumption 5-NS, while $\boldsymbol{A}$ is a stable matrix with all eigenvalues strictly inside the unit circle. The latter condition is sufficient to ensure controllability of the system. Our UC models, although having a unit root in inflation, still satisfy the weaker condition of unit-circle controllability, which together with detectability assures the existence of a steady state Kalman filter; see Section S2.2 below. After summarizing the identification conditions established by KN, we extend their framework to handle the case of unit roots.

First, build the innovations representation of the SSM of equations (A.23)

$$\boldsymbol{X}_{t+1|t+1} = \boldsymbol{A}\boldsymbol{X}_{t|t} + \boldsymbol{K}\boldsymbol{a}_{t+1} \quad \text{and} \quad \boldsymbol{Y}_{t+1} = \boldsymbol{C}\boldsymbol{X}_{t|t} + \boldsymbol{a}_{t+1}, \tag{A.24}$$

by defining expectations and innovations as $\boldsymbol{X}_{t|t} \equiv \mathbf{E}(\boldsymbol{X}_t|\boldsymbol{Y}^t)$, $\boldsymbol{Y}_{t|t-1} \equiv \mathbf{E}(\boldsymbol{Y}_t|\boldsymbol{Y}^{t-1})$ and $\boldsymbol{a}_t \equiv \boldsymbol{Y}_t - \boldsymbol{Y}_{t|t-1}$, which recreates equations (9a) and (9b) of KN, where $Y^t = \{Y_i\}_{i=1}^t$ and $\boldsymbol{K}$ is the Kalman gain to be derived next. Denote the posterior variance of the states as $\boldsymbol{\Sigma} \equiv \mathbf{Var}(\boldsymbol{X}_t|\boldsymbol{Y}^t)$ and the covariance matrix of a_t by $\boldsymbol{\Sigma}_a \equiv \mathbf{Var}(\boldsymbol{a}_t) = \mathbf{Var}(\boldsymbol{Y}_t|\boldsymbol{Y}^{t-1})$. Apply standard steady-state Kalman filtering formulas to find

$$\begin{aligned}
\boldsymbol{\Sigma}_a &= \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}' + \boldsymbol{D}\boldsymbol{D}', \\
\boldsymbol{K} &= (\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{C}' + \boldsymbol{B}\boldsymbol{D}')\boldsymbol{\Sigma}_a^{-1}, \\
\boldsymbol{\Sigma} &= \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}' - (\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{C}' + \boldsymbol{B}\boldsymbol{D}')(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}' + \boldsymbol{D}\boldsymbol{D}')^{-1}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{C}' + \boldsymbol{B}\boldsymbol{D}')'.
\end{aligned} \tag{A.25}$$

Importantly, provided that $(\boldsymbol{A}, \boldsymbol{C})$ are detectable and the system is unit-circle controllable, a stabilizing solution to the Riccati equation, which is a solution where $\boldsymbol{A} - \boldsymbol{K}\boldsymbol{C}$

[6]KN assume $\mathbf{Var}(\varepsilon_t)$ is positive definite with Choleski factor $\boldsymbol{L}_\varepsilon$, where $\mathbf{Var}(\boldsymbol{\varepsilon}_{t+1}) = \boldsymbol{L}_\varepsilon \boldsymbol{L}_\varepsilon'$. Models with a nonnormalized $\mathbf{Var}(\varepsilon_{t+1})$ can be normalized by post-multiplying $\boldsymbol{B}$ and $\boldsymbol{D}$ with $\boldsymbol{L}_\varepsilon$. The covariance matrix of the shocks can be counted as part of Ψ and their identification can be assessed as well.

[7]See their assumption 4-NS together with our normalization of $\mathbf{Var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{I}$, which is without loss of generality, because KN's assumption 1 requires $\mathbf{Var}(\varepsilon_t)$ to be positive definite.

is a stable matrix, exists and is unique; see Section S2.2 below. Notice $\boldsymbol{K}$, Σ, Σ_a and the sequence of innovations of $\boldsymbol{a}_t$ depend on the Ψ through their dependence on $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$.

Conditions for identification are given by KN's Proposition 1-NS.[8] Two parameter vectors Ψ_0 and Ψ_1 are said to be observationally equivalent if and only if there exists a full-rank $N_x \times N_x$ matrix $\boldsymbol{T}$ such that

$$\boldsymbol{A}(\Psi_1) = \boldsymbol{T}\boldsymbol{A}(\Psi_0)\boldsymbol{T}^{-1}, \qquad \boldsymbol{C}(\Psi_1) = \boldsymbol{C}(\Psi_0)\boldsymbol{T}^{-1}, \tag{A.26}$$

$$\boldsymbol{K}(\Psi_1) = \boldsymbol{T}\boldsymbol{K}(\Psi_0), \qquad \boldsymbol{\Sigma}_a(\Psi_1) = \boldsymbol{\Sigma}_a(\Psi_0). \tag{A.27}$$

When the above conditions hold, Ψ_0 and Ψ_1 generate identical innovation representations and thus identical spectral density functions for the vector of observables $\boldsymbol{Y}_t$.

Based on the innovations representation given by (A.24), and for given state space matrices $\boldsymbol{A}$, $\boldsymbol{C}$ and $\boldsymbol{K}$, and a prior for the initial state, $\boldsymbol{X}_{0|0}$, a sequence of innovations in the observables, $\boldsymbol{a}^T$, can be generated from a sequence of data $\boldsymbol{Y}^T$ using

$$\boldsymbol{a}_{t+1} = \boldsymbol{Y}_{t+1} - \boldsymbol{C}\boldsymbol{X}_{t|t} \quad \text{and} \quad \boldsymbol{X}_{t+1|t+1} = (\boldsymbol{A} - \boldsymbol{K}\boldsymbol{C})\boldsymbol{X}_{t|t} + \boldsymbol{K}\boldsymbol{Y}_{t+1}. \tag{A.28}$$

Assume for two different parameter vectors Ψ_1 and Ψ_0, that a nonsingular $\boldsymbol{T}$ exists so that (A.26) and (A.27) hold. In this case, both parametrizations are observationally equivalent because they lead to the same sequence of innovations $\boldsymbol{a}^T$ with identical variance $\boldsymbol{\Sigma}_a(\Psi_1) = \boldsymbol{\Sigma}_a(\Psi_1)$. Also, KN show both parametrizations imply the same spectral density for the observables $\boldsymbol{Y}_t$.

The analysis of KN is focused on identification via the spectral density of Y_t. Hence, their identification analysis rests on the second moments of Y_t without specific distributional assumptions. Since our framework is different, we assume that the shocks, $\boldsymbol{\varepsilon}_t$, and the priors for the initial values of the states, X_0, are normally distributed.[9] With Gaussian shocks and priors innovations in the observables are also normally distributed, $\boldsymbol{a}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_a)$. Hence, for a given sequence of data, $\boldsymbol{Y}^T$, we obtain the log likelihood as a function of the parameter vector via the usual prediction error decomposition

$$\log \mathcal{L}\big(\Psi | \boldsymbol{Y}^T; \boldsymbol{X}_{0|0}, \boldsymbol{\Sigma}_{0,0}\big) = -\frac{1}{2}\sum_{t=1}^{T}\big(N_y \cdot \log(2\pi) + \log|\boldsymbol{\Sigma}_a| + \boldsymbol{a}_t\boldsymbol{\Sigma}_a^{-1}\boldsymbol{a}_t'\big),$$

with $\boldsymbol{a}_t$ a function of Ψ as shown in (A.28). Throughout, we assume that evaluations of the likelihood are conducted for a given prior mean and variance of the initial states, $\boldsymbol{X}_{0|0}$ and $\boldsymbol{\Sigma}_{0,0}$, that are consistent with the steady state solution of the Kalman filter.

S2.2 *Some concepts from control theory*

This section reviews concepts from control theory, notably detectability and unit-circle controllability and their relevance for the existence of a steady-state Kalman filter. Readers familiar with this material may also skip this section without loss of continuity.

[8]Depending on the number of shocks and observables, KN consider both singular and nonsingular cases. In our applications, we always have $N_y \leq N_\varepsilon$, and thus belong to the nonsingular category.

[9]The analysis could also be in terms of the quasi-likelihood of a SSM with non-Gaussian shocks.

Formal conditions for the existence of a time-invariant Kalman filter have been stated, among others, by Anderson and Moore (1979), Harvey (1989), Kailath, Sayed, and Hassibi (2000), and Hansen and Sargent (2007). Necessary and sufficient conditions for the existence of a unique and stabilizing solution that is also positive semidefinite depend on the "detectability" and "unit-circle controllability" of certain matrices in our state space. We restate those concepts next.

DEFINITION (Detectability). A pair of matrices $(\boldsymbol{A}, \boldsymbol{C})$ is detectable when no right eigenvector of $\boldsymbol{A}$ that is associated with an unstable eigenvalue is orthogonal to the row space of $\boldsymbol{C}$. That is, there is no nonzero column vector $\boldsymbol{v}$ such that $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}\lambda$ and $|\lambda| \geq 1$ with $\boldsymbol{C}\boldsymbol{v} = \boldsymbol{0}$.

Detectability alone is already sufficient for the existence of *some* solution to the Riccati equation such that $\boldsymbol{A} - \boldsymbol{K}\boldsymbol{C}$ is stable; see Table E.1 in Kailath, Sayed, and Hassibi (2000). Evidently, detectability is assured when $\boldsymbol{A}$ is a stable matrix, regardless of $\boldsymbol{C}$.

DEFINITION (Unit-Circle Controllability). The pair $(\boldsymbol{A}, \boldsymbol{B})$ is unit-circle controllable when no left-eigenvector of $\boldsymbol{A}$ associated with an eigenvalue on the unit circle is orthogonal to the column space of $\boldsymbol{B}$. That is, there is no nonzero row vector $\boldsymbol{v}$ such that $\boldsymbol{v}\boldsymbol{A} = \boldsymbol{v}\lambda$ with $|\lambda| = 1$ and $\boldsymbol{v}\boldsymbol{B} = \boldsymbol{0}$.

In order to consider the role of unit-circle controllability for the existence of a stabilizing solution to the Riccati equation, it is useful to define the following two matrices[10]

$$\boldsymbol{A}^G \equiv \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}'\left(\boldsymbol{D}\boldsymbol{D}'\right)^{-1}\boldsymbol{C} \quad \text{and} \quad \boldsymbol{B}^G \equiv \boldsymbol{B}\underbrace{\left(\boldsymbol{I} - \boldsymbol{D}'\left(\boldsymbol{D}\boldsymbol{D}'\right)^{-1}\boldsymbol{D}\right)}_{\mathcal{M}^D}.$$

The concepts of detectability and unit-circle controllability provide sufficient conditions for existence and uniqueness of a stabilizing solution to the Kalman filtering problem.

THEOREM 1 (Stabilizing Solution to Riccati Equations). *Provided $\boldsymbol{D}\boldsymbol{D}'$ has full rank, a stabilizing, positive semidefinite solution to the Riccati equation* (A.25) *exists when $(\boldsymbol{A}^G, \boldsymbol{B}^G)$ is unit-circle controllable and $(\boldsymbol{A}, \boldsymbol{C})$ is detectable. The steady state Kalman gain is such that $\boldsymbol{A} - \boldsymbol{K}\boldsymbol{C}$ is a stable matrix. Moreover, the stabilizing solution is unique.*[11]

PROOF. See Theorem E.5.1 of Kailath, Sayed, and Hassibi (2000); related results are also presented in Anderson, McGrattan, Hansen, and Sargent (1996), or Chapter 4 of Anderson and Moore (1979). □

As we argue next for the case of $\boldsymbol{C} = \boldsymbol{G}\boldsymbol{A}$ and $\boldsymbol{D} = \boldsymbol{G}\boldsymbol{B}$, a sufficient condition for unit-circle controllability of $(\boldsymbol{A}^G, \boldsymbol{B}^G)$ is $\boldsymbol{B}$ has full rank. Coupled with detectability of $(\boldsymbol{A}, \boldsymbol{C})$, full rank of $\boldsymbol{B}$ ensures existence of the steady state Kalman filter. Given $\boldsymbol{C} = \boldsymbol{G}\boldsymbol{A}$

[10]These transformation are designed to handle correlation between the shocks to signal and state equations that arise when $\boldsymbol{B}\boldsymbol{D}' \neq \boldsymbol{0}$.

[11]There may be other, nonstabilizing positive semidefinite solutions.

and $\boldsymbol{D} = \boldsymbol{G}\boldsymbol{B}$, the previous expressions for $\boldsymbol{A}^G$ and $\boldsymbol{B}^G$ become $\boldsymbol{A}^G = (\boldsymbol{I} - \boldsymbol{P}^G)\boldsymbol{A}$ and $\boldsymbol{B}^G = (\boldsymbol{I} - \boldsymbol{P}^G)\boldsymbol{B}$, where $\boldsymbol{P}^G \equiv \boldsymbol{B}\boldsymbol{G}'(\boldsymbol{G}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{G}')^{-1}\boldsymbol{G} \Rightarrow \boldsymbol{P}^G = \boldsymbol{P}^G\boldsymbol{P}^G$. Hence, $\boldsymbol{P}^G$ is a nonsymmetric, idempotent projection matrix, where $\boldsymbol{G}\boldsymbol{P}^G = \boldsymbol{G}$.[12] In addition, when $\boldsymbol{C} = \boldsymbol{G}\boldsymbol{A}$ and $\boldsymbol{D} = \boldsymbol{G}\boldsymbol{B}$, unit-circle controllability of $(\boldsymbol{A}^G, \boldsymbol{B}^G)$ is equivalent to unit-circle controllability of $(\boldsymbol{A}(\boldsymbol{I} - \boldsymbol{P}^G), \boldsymbol{B})$. Define $\tilde{\boldsymbol{v}} \equiv \boldsymbol{v}(\boldsymbol{I} - \boldsymbol{P}^G)$ and note left-eigenvectors of $\boldsymbol{A}^G$ associated with eigenvalues on the unit circle cannot be orthogonal to $\boldsymbol{P}^G$. Otherwise, $\boldsymbol{v}\boldsymbol{A}^G = \boldsymbol{0}$. Hence, for $|\lambda| = 1$, $\boldsymbol{v}\boldsymbol{A}^G = \boldsymbol{v}\lambda$, $\boldsymbol{v}\boldsymbol{B}^G \neq \boldsymbol{0}$ and $\boldsymbol{v} \neq \boldsymbol{0}$ are equivalent to $\tilde{\boldsymbol{v}}\boldsymbol{A}(\boldsymbol{I} - \boldsymbol{P}^G) = \tilde{\boldsymbol{v}}\lambda$, $\tilde{\boldsymbol{v}} \neq \boldsymbol{0}$, and $\tilde{\boldsymbol{v}}\boldsymbol{B} \neq \boldsymbol{0}$. A sufficient condition for these equivalences is $\boldsymbol{B}$ has full rank.

S2.3 *Extension of KN to observables with unit root dynamics*

The analysis of KN is based on identification via the spectral density of $\boldsymbol{Y}_t$ which entails the requirement that $\boldsymbol{Y}_t$ is stationary. In addition, KN also require the state vector $\boldsymbol{X}_t$ to be stationary. However, our UC models typically have a unit root in the state dynamics which is inherited by the observables as well.

We add the assumption that the shock vector has a standard normal distribution, but relax the original KN assumption that $\boldsymbol{A}(\Psi)$ is a stable matrix. Also, our notion of observational equivalence is in terms of whether two different parameter vector, Ψ_1 and Ψ_2, generate different log likelihoods for a given sequence of data $\boldsymbol{Y}^T$. Given a positive definite solution exists to the Riccati equations (A.25), the KN conditions (A.26) and (A.27) remain necessary and sufficient to establish whether two different parametrizations, Ψ_1 and Ψ_2, yield the same likelihoods, $\log \mathcal{L}(\Psi_1|\boldsymbol{Y}^T) = \log \mathcal{L}(\Psi_0|\boldsymbol{Y}^T)$.

The KN assumption of a stable state transition matrix $\boldsymbol{A}$, coupled with a positive definite $\textbf{Var}(\boldsymbol{\varepsilon}_t)$ and $(\boldsymbol{A}, \boldsymbol{C})$ detectable (see KN assumptions 1 and 5-NS), is sufficient to ensure existence of a positive-definite solution $\boldsymbol{\Sigma}$ to the Riccati equations (A.25).[13] However, existence of a solution to the Riccati equations can also be guaranteed by requiring simply that $(\boldsymbol{A}, \boldsymbol{B})$ is unit-circle controllable while maintaining detectability of $(\boldsymbol{A}, \boldsymbol{C})$ and strict positive-definiteness of $\boldsymbol{D}\boldsymbol{D}'$; see our Appendix for further details. As long as only parametrizations are considered for which $\boldsymbol{D}\boldsymbol{D}'$ has full rank, $(\boldsymbol{A}, \boldsymbol{C})$ is detectable, and $(\boldsymbol{A}, \boldsymbol{B})$ is unit-circle controllable, the KN conditions (A.26) and (A.27) continue to characterize observationally equivalent parametrizations. Below, we will establish that in the class of UC models considered in our paper, these conditions can only hold when the two parameter vectors are identical, $\Psi_1 = \Psi_2$.

As in KN, our discussion has been based on Kalman filtering recursions that use steady-state values for the Kalman gain. This is consistent with a prior distribution over $\boldsymbol{X}_0$ that is identical to the ergodic distribution of $\boldsymbol{X}_t$. However, when allowing for unit-root (or other nonstationary) dynamics in the state vector, such an ergodic distribution does not exist (at least not for the nonstationary elements of the state vector). Given the conditions of detectability and unit-circle controllability discussed above hold, a steady state solution to the Riccati equation, and thus also for the Kalman gain, continue to exist. As noted above, to avoid consideration of initial heteroscedasticity in the posterior variances, we assume the Kalman filter is initialized with a prior variance matrix, $\boldsymbol{\Sigma}_{0|0}$ that is consistent with the filter's steady state solution.

[12] The eigenvalues of an idempotent matrix are either zero or one. In this case, $|\boldsymbol{P}^G| = \boldsymbol{0}$.

[13] Recall that we use the standardization $\textbf{Var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{I}$, which is positive definite.

S2.4 *KN applied to the univariate Stock–Watson UC model*

Before considering our SSM for the joint dynamics of inflation and SI forecasts, this section illustrates the use of the KN framework to establish identification in SW-UC models for inflation with or without persistence in the inflation gap. First, consider the original SW-UC model with a univariate measurement equation, but without the irregular gap component that was introduced in equation (A.6). This model maps into the KN framework as $\boldsymbol{X}_t = [\tau_t \ \varepsilon_t]'$, $\boldsymbol{Y}_t = \pi_t$, $\boldsymbol{A} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \theta \end{smallmatrix}\right]$, $\boldsymbol{B} = \left[\begin{smallmatrix} \varsigma_\eta & 0 \\ 0 & \varsigma_\upsilon \end{smallmatrix}\right]$, $\boldsymbol{C} = [1 \ \theta]$, and $\boldsymbol{D} = [\varsigma_\eta \ \varsigma_\upsilon]$, where $\Psi = [\theta \ \varsigma_\eta^2 \ \varsigma_\upsilon^2]'$, $\boldsymbol{C} = \boldsymbol{G}\boldsymbol{A}$, and $\boldsymbol{G} = [1 \ 1]$.[14]

For now, restrict attention to the case $\theta \neq 0$, $|\boldsymbol{A}| \neq 0$. Also, maintain the assumption that θ lies inside the unit circle. We will return to the case of $\theta = 0$ in our discussion of the UC model with an irregular gap component that is serially uncorrelated. It is straightforward to show the only nonsingular choice for $\boldsymbol{T}$ for which the KN conditions apply is the 2×2 identity matrix. The upshot is two observationally equivalent parameter vectors, Ψ_1 and Ψ_0, have to agree on the value for the scalar parameter θ to ensure $\boldsymbol{A}(\Psi_1) = \boldsymbol{A}(\Psi_0)$ and $\boldsymbol{C}(\Psi_1) = \boldsymbol{C}(\Psi_0)$. Given these equalities, henceforth, we refer to these matrices as $\boldsymbol{A}$ and $\boldsymbol{C}$.[15]

With $\boldsymbol{T} = \boldsymbol{I}$, the KN conditions imply that both parameter vectors should generate the same covariance between innovations in observables and latent states as well as the same variance for projections of the latent states onto the observables. To see this, denote $\text{Cov}\,(\boldsymbol{X}_t, \boldsymbol{Y}_t | \boldsymbol{Y}^{t-1})$ by $\boldsymbol{\Sigma}_{XY}$ and $\textbf{Var}(\boldsymbol{X}_{t|t} | \boldsymbol{Y}^{t-1})$ by $\hat{\boldsymbol{\Sigma}}$. Furthermore, note $\boldsymbol{K} = \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_a^{-1}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{K} \boldsymbol{\Sigma}_a \boldsymbol{K}'$. If Ψ_1 and Ψ_2 are to satisfy the KN conditions, which include $\boldsymbol{\Sigma}_a(\Psi_1) = \boldsymbol{\Sigma}_a(\Psi_1)$; it follows $\boldsymbol{\Sigma}_{XY}(\Psi_1) = \boldsymbol{\Sigma}_{XY}(\Psi_0)$ and $\hat{\boldsymbol{\Sigma}}(\Psi_1) = \hat{\boldsymbol{\Sigma}}(\Psi_0)$ are true.

Taking differences between the Riccati equations (A.25) generated by both parameter vectors, we obtain $\boldsymbol{\Delta} \equiv \boldsymbol{\Sigma}(\Psi_1) - \boldsymbol{\Sigma}(\Psi_0) = \boldsymbol{A} \boldsymbol{\Delta} \boldsymbol{A}' + \boldsymbol{B}(\boldsymbol{\theta}_1) \boldsymbol{B}(\boldsymbol{\theta}_1)' - \boldsymbol{B}(\boldsymbol{\theta}_0) \boldsymbol{B}(\boldsymbol{\theta}_0)'$. Since $\boldsymbol{A}$ along with $\boldsymbol{B_1}$ and $\boldsymbol{B_0}$ are diagonal, $\boldsymbol{\Delta}$ must be diagonal $\boldsymbol{\Delta} = \left[\begin{smallmatrix} \Delta_{11} & 0 \\ 0 & \Delta_{22} \end{smallmatrix}\right]$, where $\Delta_{11} = \Delta_{11} + \sigma_{\eta,1}^2 - \sigma_{\eta,0}^2$ and $\Delta_{22} = \theta^2 \Delta_{22} + \sigma_{\upsilon,1}^2 - \sigma_{\upsilon,0}^2$.[16] Given $\boldsymbol{G} = [1 \ 1]$, $\boldsymbol{Y}_t = \boldsymbol{G}\boldsymbol{X}_t$ and $\boldsymbol{\Sigma}_{XY} = (\boldsymbol{\Sigma} + \hat{\boldsymbol{\Sigma}}) \boldsymbol{G}'$. Hence, $\boldsymbol{\Sigma}_{XY}(\Psi_1) = \boldsymbol{\Sigma}_{XY}(\Psi_0)$ and $\hat{\boldsymbol{\Sigma}}(\Psi_1) = \hat{\boldsymbol{\Sigma}}(\Psi_0)$, which yields $\boldsymbol{\Delta}\boldsymbol{G}' = [\Delta_{11} \ \Delta_{22}]' = \boldsymbol{0}$, $\sigma_{\eta,1}^2 = \sigma_{\eta,0}^2$, and $\sigma_{\upsilon,1}^2 = \sigma_{\upsilon,0}^2$. We conclude $\Psi_1 = \Psi_0$.

S2.4.1 *Stock–Watson UC model with irregular gap component* Before studying the case $\theta = 0$, we consider an augmented UC model with $\theta \neq 0$ and irregular gap component, ζ_t, in inflation. The measurement equation is $\pi_t = \tau_t + \varepsilon_t + \zeta_t$, where $\zeta_t \sim N(0, \varsigma_\zeta^2)$. We are able to maintain the definitions of the state space matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ used in the previous section while amending the definition of $\boldsymbol{D}$ to $\boldsymbol{D} = [\varsigma_\eta \ \varsigma_\upsilon \ \varsigma_\zeta]$. As before, the conclusions are that only cases with $\boldsymbol{T} = \boldsymbol{I}$ need to be considered and the values of θ

[14] The conditions for existence of a steady state Kalman filter, detectability and unit-circle controllability as defined in Section S2.2, are met. $(\boldsymbol{A}, \boldsymbol{C})$ are detectable and $\boldsymbol{B}$ has full rank.

[15] The specific structure of $\boldsymbol{A}$ and $\boldsymbol{C}$ in this UC model requires $\boldsymbol{T} = \boldsymbol{I}$. Let $\boldsymbol{1}$ denote a vector of ones. Together with $\boldsymbol{C} = \boldsymbol{1}\boldsymbol{A}$, the KN conditions stated in (A.26) require $\boldsymbol{1}(\boldsymbol{A}(\Psi_1)\boldsymbol{T} - \boldsymbol{A}(\Psi_0)) = 0 \Leftrightarrow \boldsymbol{1}(\boldsymbol{T} - \boldsymbol{I})\boldsymbol{A}(\Psi_0) = 0$ because $\boldsymbol{A}$ is nonsingular. As a result, the only nonsingular choice of $\boldsymbol{T}$ for (A.26) to hold is $\boldsymbol{T} = \boldsymbol{I}$. Alternatively, note $\boldsymbol{A}(\Psi_1) = \boldsymbol{T}\boldsymbol{A}(\Psi_0)\boldsymbol{T}^{-1}$. This requires $\boldsymbol{T}$ to be diagonal. The reason is $\boldsymbol{A}$ is diagonal, which requires the AR parameter $\theta_1 = \theta_0$ to be identical under both parametrizations. As a result, $\boldsymbol{C}(\Psi_1) = \boldsymbol{C}(\Psi_0)\boldsymbol{T}^{-1}$ restricts the diagonal elements of $\boldsymbol{T}$ to unit values.

[16] Note that $\Delta_{11} = \Delta_{11} + \sigma_{\eta,1}^2 - \sigma_{\eta,0}^2$ already requires $\sigma_{\eta,1}^2 = \sigma_{\eta,0}^2$.

must be identical under both candidate parametrizations. Although $\boldsymbol{Y}_t = \boldsymbol{G}\boldsymbol{X}_t + \zeta_t$, we still conclude $\boldsymbol{\Delta G}' = 0$, and thus $\sigma^2_{\eta,1} = \sigma^2_{\eta,0}$ and $\sigma^2_{\upsilon,1} = \sigma^2_{\upsilon,0}$.[17]

The only remaining parameter to be identified is ς^2_ζ. The KN conditions require $\boldsymbol{\Sigma}_a(\Psi_1) = \boldsymbol{\Sigma}_a(\Psi_0)$, $\hat{\boldsymbol{\Sigma}}(\Psi_1) = \hat{\boldsymbol{\Sigma}}(\Psi_0)$, and $\boldsymbol{G}\boldsymbol{\Sigma}(\Psi_1) = \boldsymbol{G}\boldsymbol{\Sigma}(\Psi_0)$. Since, $\boldsymbol{\Sigma}_a = \boldsymbol{G}(\boldsymbol{\Sigma} + \hat{\boldsymbol{\Sigma}})\boldsymbol{G}' + \varsigma^2_\zeta$, it follows Ψ_1 and Ψ_0 must have the same element for ς^2_ζ. The original UC model of Stock and Watson (2007), which had *i.i.d.* gap inflation, is nested by this more general case.

S2.5 *Identification in a constant-parameter version of our SSMs*

This section considers identification in a constant-parameter version of our SSMs, which we will call the "SI-UC" model. This constant-parameter corresponds to the SSM $\mathcal{M}_1$, where the AR(1) coefficients θ and the SI parameter λ were constant, with the added restrictions that the variances of shocks to trend and gap inflation are also constant. The only remaining latent variables are the linear states of the model, which are the RE and SI versions of trend and gap inflation.

For simplicity, we restrict attention to the case where a survey forecast at an arbitrary horizon h is observed in addition to realized inflation. Noisy observations of forecasts at a single horizon are already sufficient to identify the additional parameters of the SI-UC model compared with the SW-UC model. Adding forecasts at other horizons to the measurement vector adds only the measurement variances associated with those additional measurements to the parameter vector, which are easily identified.

Before establishing the algebraic arguments in details, we note the inflation process is unaffected by the survey data block of the model. This suggests identification of parameters of the inflation process should work as in the UC model. The only additional parameters to be identified by the measurement equations for surveys are the SI weight, λ, and the measurement error associated with each survey.

The SI-UC model has the following measurement, state, and shock vectors $\boldsymbol{Y}_t = [\pi_t \; \mathbf{F}_t\pi_{t+h} + \zeta^h_t]'$, $\boldsymbol{X}_t = [\tau_t \; \varepsilon_t \; \mathbf{F}_t\tau_t \; \mathbf{F}_t\varepsilon_t]'$, and $\boldsymbol{\varepsilon}_t = [\eta_t \; \upsilon_t \; \zeta^\pi_t \; \zeta^h_t]'$. Assuming constant AR1 inflation gap persistence and SI weight parameters and shock variances, the ABCD state space matrices are

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \theta & 0 & 0 \\ 1-\lambda & 0 & \lambda & 0 \\ 0 & (1-\lambda)\theta & 0 & \lambda\theta \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} \varsigma_\eta & 0 & 0 & 0 \\ 0 & \varsigma_\upsilon & 0 & 0 \\ (1-\lambda)\varsigma_\eta & 0 & 0 & 0 \\ 0 & (1-\lambda)\varsigma_\upsilon & 0 & 0 \end{bmatrix},$$

$\boldsymbol{C} = \begin{bmatrix} 1 & \theta & 0 & 0 \\ 0 & 0 & 1 & \theta^{h+1} \end{bmatrix}$, and $\boldsymbol{D} = \begin{bmatrix} \varsigma_\eta & \varsigma_\upsilon & \varsigma_{\zeta,\pi} & 0 \\ (1-\lambda)\varsigma_\eta & (1-\lambda)\varsigma_\upsilon & 0 & \varsigma_{\zeta,h} \end{bmatrix}$.[18] The parameter vector is $\Psi = [\theta \; \lambda \; \varsigma^2_\eta \; \varsigma^2_\upsilon \; \varsigma^2_{\zeta,\pi} \; \varsigma^2_{\zeta,h}]'$.

[17] It is easy to verify the conditions for existence of a steady state Kalman filter are met. Augment the state vector by the irregular gap component, which does not add unit roots to the transition matrix $\boldsymbol{A}$. The rewritten model is $\boldsymbol{Y}_t = \boldsymbol{G}\boldsymbol{X}_t$, where $\boldsymbol{B}$ has full rank, which ensures that the existence conditions are met, as reviewed in Section S2.2.

[18] As before, verifying that conditions for the existence of a steady state Kalman filter are met is straightforward, since the model has the form $\boldsymbol{Y}_t = \boldsymbol{G}\boldsymbol{X}_t$ with a $\boldsymbol{B}$ matrix that has full rank.

As before, we seek to establish that the KN conditions (A.26) and (A.27) are satisfied only if two candidate parametrizations coincide, $\Psi_1 = \Psi_0$. The analysis is facilitated by noting (A.26) implies the KN conditions can only hold for $\boldsymbol{T} = \boldsymbol{I}$. For (A.26) to hold, both candidate parametrizations must have the same values for λ and θ. To establish the KN conditions hold only if $\boldsymbol{T} = 1$, the state space matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ involved in the first two KN conditions given by (A.26), are $\boldsymbol{A} = \left[\begin{smallmatrix} \boldsymbol{\Theta} & \mathbf{0} \\ (1-\lambda)\boldsymbol{\Theta} & \lambda\boldsymbol{\Theta} \end{smallmatrix}\right]$, $\boldsymbol{\Theta} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \theta \end{smallmatrix}\right]$, $\boldsymbol{C} = \left[\begin{smallmatrix} \boldsymbol{G\Theta} & \mathbf{0} \\ \mathbf{0} & \lambda\boldsymbol{G\Theta}^{h+1} \end{smallmatrix}\right]$, and $\boldsymbol{G} = [1\ 1]$. Next, partition $\boldsymbol{T}$ into four 2×2 matrices

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix}. \tag{A.29}$$

The result is the first two KN conditions, given in (A.26), must be

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{\Theta}(\Psi_1)\boldsymbol{T}_{11} & \boldsymbol{\Theta}(\Psi_1)\boldsymbol{T}_{12} \\ \boldsymbol{\Theta}(\Psi_1)\big((1-\lambda_1)\boldsymbol{T}_{11} + \lambda_1\boldsymbol{T}_{21}\big) & \boldsymbol{\Theta}(\Psi_1)\big((1-\lambda_1)\boldsymbol{T}_{12} + \lambda_1\boldsymbol{T}_{22}\big) \end{bmatrix} \\ &\quad = \begin{bmatrix} \big(\boldsymbol{T}_{11} + (1-\lambda_0)\boldsymbol{T}_{12}\big)\boldsymbol{\Theta}(\Psi_0) & \lambda_0\boldsymbol{T}_{12}\boldsymbol{\Theta}(\Psi_0) \\ \big(\boldsymbol{T}_{21} + (1-\lambda_0)\boldsymbol{T}_{22}\big)\boldsymbol{\Theta}(\Psi_0) & \lambda_0\boldsymbol{T}_{22}\boldsymbol{\Theta}(\Psi_0) \end{bmatrix}, \end{aligned} \tag{A.30}$$

and

$$\begin{bmatrix} \boldsymbol{G\Theta}(\Psi_1)\boldsymbol{T}_{11} & \boldsymbol{G\Theta}(\Psi_1)\boldsymbol{T}_{12} \\ \lambda_1\boldsymbol{G\Theta}(\Psi_1)^{h+1}\boldsymbol{T}_{21} & \lambda_1\boldsymbol{G\Theta}(\Psi_1)^{h+1}\boldsymbol{T}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G\Theta}(\Psi_0) & \mathbf{0} \\ \mathbf{0} & \lambda_0\boldsymbol{G\Theta}(\Psi_0)^{h+1} \end{bmatrix}, \tag{A.31}$$

where λ_0 and λ_1 denote the entries for λ contained in Ψ_0 and Ψ_0, respectively. Notice $\boldsymbol{\Theta}$ is the 2×2 transition matrix of the SW-UC model whose identification has been established above. For the SW-UC model, the KN conditions require $\boldsymbol{\Theta}(\Psi_1)\boldsymbol{T}_{11} = \boldsymbol{T}_{11}\boldsymbol{\Theta}(\Psi_0)$ and $\boldsymbol{G\Theta}(\Psi_1)\boldsymbol{T}_{11} = \boldsymbol{G\Theta}(\Psi_0)$ for some invertible square matrix $\boldsymbol{T}_{11}$. This led us to conclude that $\boldsymbol{T}_{11} = \boldsymbol{I}$ and $\boldsymbol{\Theta}(\Psi_1) = \boldsymbol{\Theta}(\Psi_0)$.

Next, we show these results extend to the SI-UC model with $\boldsymbol{T}_{11}$ being the top-left partition of $\boldsymbol{T}$ as defined in (A.29). First, the top-left partitions of (A.30) and (A.31) require $\boldsymbol{T}_{12} = \mathbf{0}$. To see this, note these two conditions combined require $\boldsymbol{G\Theta}(\Psi_1)\boldsymbol{T}_{12} = \mathbf{0}$ and $\boldsymbol{GT}_{12} = \mathbf{0}$. Recall $\boldsymbol{T}_{12}$ is a 2×2 matrix and note that $\boldsymbol{G\Theta}(\Psi_1)$ and $\boldsymbol{G}$ are two non-collinear row vectors such that the null-space of $\boldsymbol{T}_{12}$ must have a full rank of two. As a result, we must have $\boldsymbol{T}_{12} = \mathbf{0}$ for the top-left partitions of (A.30) and (A.31) to hold.

Given $\boldsymbol{T}_{12} = \mathbf{0}$, the top-right partitions of (A.30) and (A.31) reduce to results known for the SW-UC model. These are $\boldsymbol{T}_{11} = \boldsymbol{I}$, $\boldsymbol{\Theta}(\Psi_1) = \boldsymbol{\Theta}(\Psi_0)$, and thus $\theta_1 = \theta_0$. These results reflect the block-triangular structure of the SI-UC model in which the inflation process is identical to the SW-UC model and unaffected by the SI forecasting block. With $\boldsymbol{T}_{11} = \boldsymbol{I}$, $\boldsymbol{T}_{12} = \mathbf{0}$ and $\boldsymbol{\Theta}(\Psi_1) = \boldsymbol{\Theta}(\Psi_0) = \boldsymbol{\Theta}$, the bottom-right partitions of (A.30) and (A.31) can only hold if $\boldsymbol{T}_{22} = \boldsymbol{I}$ and $\lambda_1 = \lambda_0$.[19] The bottom-left partitions of (A.30) and (A.31) set $\boldsymbol{T}_{12} = \mathbf{0}$. Having established that $\boldsymbol{T} = 1$ and $\theta_1 = \theta_0$, follow the steps used above for the

[19]Notice that with $\boldsymbol{T}_{12} = \mathbf{0}$, $\boldsymbol{T}$ can only be nonsingular if $\boldsymbol{T}_{22}$ is nonsingular which rules out $\boldsymbol{T}_{22} = \mathbf{0}$.

SW-UC model to show the rest of the KN conditions impose equality on the remaining variance parameters contained in Ψ_1 and Ψ_0, which gives $\Psi_1 = \Psi_0$.[20]

S3. Econometric methods

This section gives details about the sequential Monte Carlo (SMC) methods sketched in Section 3 of the paper. We employ a particle learning filter to jointly estimate the parameters of the state vector $\mathcal{T}_t$, which consists of linear state variables, $\mathcal{S}_t$, and nonlinear states $\mathcal{V}_t$, and the parameters Ψ of our SSMs. In addition to filtered estimates, we also produce smoothed estimates of $\mathcal{T}_t$ as described below.

Our particle learning filter builds on procedures described by Carvalho et al. (2010) and Lopes and Tsay (2011). Carvalho et al. (2010) developed a particle learning filter to estimate Ψ that combines Rao–Blackwellization of $\mathcal{S}_t$ with filtering of $\mathcal{V}_t$ assigned to a simulation estimator. We implement a particle learning filter using Algorithm 7 of Lopes and Tsay (2011) augmented by their Algorithm 2 that employs the auxiliary particle filtering of Pitt and Shephard (1999, 2001).

Particle learning and Rao–Blackwellization engage sufficient statistics to track conditional distributions. Considering Ψ, the sufficient statistics for the ith particle are denoted $\Gamma_t^{(i)}$ with further details provided below. Most of the SSM parameters are characterized by $\mathcal{IG}$ conjugate priors and the corresponding sufficient statistics are shape and scale parameters of the $\mathcal{IG}$ priors and resulting posteriors. Rao–Blackwellization employs Kalman filtering operations to track the normal posteriors of $\mathcal{S}_t$ conditional on Ψ and $\mathcal{V}_t$, where the sufficient statistics are the mean vector $\mathcal{S}_{t|t}^{(i)} = \mathbf{E}(\mathcal{S}_t|\mathcal{Y}^t, \mathcal{V}^{t,(i)}, \Psi^{(i)})$ and the covariance matrix $\boldsymbol{\Sigma}_{t|t}^{(i)} = \mathbf{Var}(\mathcal{S}_t|\mathcal{Y}^t, \mathcal{V}^{t,(i)}, \Psi^{(i)})$.[21]

S3.1 *Particle filter for the baseline model, $\mathcal{M}_0$*

This section applies our particle learning filter algorithm to the baseline SSM, $\mathcal{M}_0$, with time-varying SI weight, λ_t, but constant θ, has seven state variables. These states are grouped into $\mathcal{V}_t = [\varsigma_{\eta,t}\ \varsigma_{\upsilon,t}\ \lambda_t]'$, $\mathcal{S}_t = [\tau_t\ \varepsilon_t\ \mathbf{F}_t\tau_t\ \mathbf{F}_t\varepsilon_t]'$. The parameter vector of $\mathcal{M}_0$ consists of the variances of shocks to the nonlinear states, variance of the irregular inflation gap component, the measurement error variances, and θ, $\Psi_0 = [\sigma_\eta^2\ \sigma_\upsilon^2\ \sigma_\kappa^2\ \sigma_{\zeta,\pi}^2\ \sigma_{\zeta,1}^2\ \sigma_{\zeta,2}^2\ \cdots\ \sigma_{\zeta,\mathcal{H}}^2\ \theta]'$.

S3.1.1 *Initialization of particle filter* For M particles indexed by i, and $t = 0$,

1. set sufficient statistics $\Gamma_0^{(i)}$ using each parameter's prior described in Section 4 of the paper. For θ, the prior is a truncated normal, as described in Table 3 of the paper, and sufficient statistics are mean and variance. The remaining elements of Ψ_0 have $\mathcal{IG}$ priors characterized by shape and scale parameters α and β with prior values listed in Table 2 of the paper.

[20]The SI-UC measurement equation is written $\boldsymbol{Y}_t = \bar{\boldsymbol{G}}\boldsymbol{X}_t + \boldsymbol{\zeta}_t$, where $\bar{G} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & \theta^h \end{bmatrix}$ and $\zeta_t = [\,\zeta_{\pi,t}\ \zeta_{h,t}\,]'$ contains the irregular inflation gap component and the measurement error in the survey observation.

[21]Kalman filtering recursions for $\mathcal{S}_{t|t}^{(i)}$ and $\boldsymbol{\Sigma}_{t|t}^{(i)}$ are given by (A.32) and (A.33) below.

2. Draw initial values for the parameters $\Psi_0^{(i)} \sim p(\Psi|\Gamma_0^{(i)})$,
3. initial values for the nonlinear states, $\mathcal{V}_0^{(i)}$, from priors described in Table 4 of the paper,
4. and $\mathcal{S}_{0|0}^{(i)}$ and $\boldsymbol{\Sigma}_{0|0}^{(i)}$ from the normal priors for the linear states described in Table 4 of the paper.[22]
5. Thus, the initial swarm of particles consists of $\{\mathcal{V}_0^{(i)}, \Gamma_0^{(i)}, \Psi_0^{(i)}, \mathcal{S}_{0|0}^{(i)}, \boldsymbol{\Sigma}_{0|0}^{(i)}\}_{i=1}^M$ and are weighted by $W_0^{(i)} = 1/M$, where it is understood that the zero subscript on Ψ_0 refers to $\mathcal{M}_0$ and not to date $t = 0$.

S3.1.2 *Particle filter iterations* Given $\{\mathcal{V}_{t-1}^{(i)}, \Gamma_{t-1}^{(i)}, \Psi_0^{(i)}, \mathcal{S}_{t-1|t-1}^{(i)}, \boldsymbol{\Sigma}_{t-1|t-1}^{(i)} W_t^{(i)}\}_{i=1}^M$, at every date $t = 1, \ldots, T$,

1. adapt the $i = 1, \ldots, M$ particles of date $t-1$ to $\mathcal{Y}_t$ by
   (a) computing the log likelihood, $\hat{\ell}_t^{(i)} \equiv \log\left(p(\mathcal{Y}_t|\mathcal{S}_{t-1|t-1}^{(i)}, \boldsymbol{\Sigma}_{t-1|t-1}^{(i)}, \widehat{\mathcal{V}}_t^{(i)}, \Psi_0^{(i)})\right)$, as described below in the Kalman filter's prediction step of Section S3.1.3,
   (b) calculating the auxiliary particle weights $W_{t-1|t}^{(i)} = w_{t-1|t}^{(i)} / \sum_{i=1}^M w_{t-1|t}^{(i)}$, where $w_{t-1|t}^{(i)} = \exp\{\hat{\ell}_t^{(i)}\} \cdot W_{t-1}^{(i)}$,
   (c) and resampling the date $t-1$ particles and sufficient statistics, $\mathcal{V}_{t-1}^{(i)}$, $\Gamma_{t-1}^{(i)}$, $\Psi^{(i)}$, $\mathcal{S}_{t-1|t-1}^{(i)}$, and $\boldsymbol{\Sigma}_{t-1|t-1}^{(i)}$ using systematic resampling from the *p.d.f.* given by $\{W_{t-1|t}^{(i)}\}_{i=1}^M$ and for use below, also resample the log likelihoods $\{\hat{\ell}_t^{(i)}\}_{i=1}^M$,
   (d) where the proposal $\widehat{\mathcal{V}}_t^{(i)} = \mathcal{V}_{t-1}^{(i)}$ is the median of the density $p(\mathcal{V}_t|V_{t-1}^{(i)}), \mathcal{V}_{t-1}^{(i)}$.
2. For every particle $i = 1, \ldots, M$,
   (a) propagate the nonlinear states by drawing $\mathcal{V}_t^{(i)} \sim p(\mathcal{V}_t|\mathcal{V}_{t-1}^{(i)}, \Psi_0^{(i)})$,
   (b) update the sufficient statistics for $\mathcal{S}_t$, $\mathcal{S}_{t|t}^{(i)} = \mathbf{E}(\mathcal{S}_t|\mathcal{Y}_t, \mathcal{S}_{t-1|t-1}^{(i)}, \boldsymbol{\Sigma}_{t-1|t-1}^{(i)}, \mathcal{V}_t^{(i)}, \Psi_0^{(i)})$ and $\boldsymbol{\Sigma}_{t|t}^{(i)} = \mathbf{Var}(\mathcal{S}_t|\mathcal{Y}_t, \mathcal{S}_{t-1|t-1}^{(i)}, \boldsymbol{\Sigma}_{t-1|t-1}^{(i)}, \mathcal{V}_t^{(i)}, \Psi_0^{(i)})$, using the Kalman filtering steps outlined in Section S3.1.3 below,
   (c) record the log likelihood, $\ell_t^{(i)}$, implied by the Kalman filter's prediction step as discussed in Section S3.1.3 below,
   (d) draw $\mathcal{S}_t^{(i)} \sim \mathcal{N}(\mathcal{S}_t|\mathcal{S}_{t|t}^{(i)}, \boldsymbol{\Sigma}_{t|t}^{(i)})$ to be used for updating sufficient statistics of Ψ in the next steps, as well as for reporting results about the posterior of $\mathcal{S}_{t|t}$,
   (e) update the sufficient statistics of θ by drawing ε_t as well as its lagged value by augmenting $\mathcal{S}_t$ to include ε_{t-1}, so that draws of $\varepsilon_{t-1}^{(i)}$ are found in $\mathcal{S}_t^{(i)}$,
   (f) update the sufficient statistics $\Gamma_t^{(i)} = \mathcal{G}(\Gamma_{t-1}^{(i)}, \mathcal{V}_t^{(i)}, \mathcal{V}_{t-1}^{(i)}, \mathcal{S}_t^{(i)}, \mathcal{Y}_t)$ with details provided in Section S3.1.4 below,

[22]The priors of ε_0 and $\mathbf{F}_0\varepsilon_0$ depend on the initial draws for $\mathcal{V}_0^{(i)}$ and $\Psi_0^{(i)}$, $i = 1, \ldots, M$.

3. compute $W_t^{(i)} = w_t^{(i)} / \sum_{i=1}^M w_t^{(i)}$, where $w_t^{(i)} = \exp\{\ell_t^{(i)}\} / \exp\{\hat{\ell}_t^{(i)}\}$, and $\hat{\ell}_t^{(i)}$ is resampled log likelihood of the auxiliary adaptation step described in step 1c above,
4. record date t moments of interest, for example, $\mathcal{V}_{t|t} = \mathbf{E}(\mathcal{V}_t|\mathcal{Y}^t) = \sum_{i=1}^M W_t^{(i)} \mathcal{V}_t^{(i)}$ and $\mathcal{S}_{t|t} = \mathbf{E}(\mathcal{S}_t|\mathcal{Y}^t) = \sum_{i=1}^M W_t^{(i)} \mathcal{S}_{t|t}^{(i)}$, and
5. generate posterior quantiles of $\mathcal{S}_t$ by simulating from a mixture of normals implied by $\{w_t^{(i)}\}_{i=1}^M$ and $\{\mathcal{S}_{t|t}, \boldsymbol{\Sigma}_{t|t}^{(i)}\}_{i=1}^M$ using the marginal data density (MDD) estimator of Pitt, dos Santos Silva, Giordani, and Kohn (2012), $p(\mathcal{Y}_t|\mathcal{Y}^{t-1}) = (M^{-1} \sum_{i=1}^M w_t^{(i)}) \cdot (\sum_{j=1}^M w_{t-1|t}^{(j)})$.

We employ the MDD estimator to sample from the posterior distribution of $\mathcal{S}_t$ because Rao–Blackwellization lowers Monte Carlo error by averaging over $\mathcal{S}_{t|t}^{(i)}$ compared with $\mathcal{S}_t^{(i)}$.

S3.1.3 *Details of the Kalman filtering steps* The Kalman filtering steps referred to in Section S3.1.2 above are as follows:

$$\mathcal{S}_{t|t-1}^{(i)} = \boldsymbol{\mathcal{A}}_t^{(i)} \mathcal{S}_{t-1|t-1}^{(i)}, \tag{A.32}$$

$$\boldsymbol{\Sigma}_{t|t-1}^{(i)} = \boldsymbol{\mathcal{A}}_t^{(i)} \boldsymbol{\Sigma}_{t-1|t-1}^{(i)} (\boldsymbol{\mathcal{A}}_t^{(i)})' + \boldsymbol{\mathcal{B}}_t^{(i)} (\boldsymbol{\mathcal{B}}_t^{(i)})', \tag{A.33}$$

$$\begin{aligned}
\boldsymbol{\Omega}_{t|t-1}^{(i)} &= \boldsymbol{\mathcal{C}}_t^{(i)} \boldsymbol{\Sigma}_{t|t-1}^{(i)} (\boldsymbol{\mathcal{C}}_t^{(i)})' + \boldsymbol{\Omega}_{\mathcal{U}}^{(i)}, \\
\widetilde{\mathcal{Y}}_t^{(i)} &= \mathcal{Y}_t - \boldsymbol{\mathcal{C}}_t^{(i)} \mathcal{S}_{t|t-1}^{(i)}, \\
\boldsymbol{\mathcal{K}}_t^{(i)} &= \boldsymbol{\Sigma}_{t|t-1}^{(i)} (\boldsymbol{\mathcal{C}}_t^{(i)})' (\boldsymbol{\Omega}_{t|t-1}^{(i)})^{-1}, \\
\mathcal{S}_{t|t}^{(i)} &= \boldsymbol{\mathcal{A}}_t^{(i)} \mathcal{S}_{t|t-1}^{(i)} + \boldsymbol{\mathcal{K}}_t^{(i)} \widetilde{\mathcal{Y}}_t^{(i)}, \\
\boldsymbol{\Sigma}_{t|t}^{(i)} &= \boldsymbol{\Sigma}_{t|t-1}^{(i)} - \boldsymbol{\Sigma}_{t|t-1}^{(i)} (\boldsymbol{\mathcal{C}}_t^{(i)})' (\boldsymbol{\Omega}_{t|t-1}^{(i)})^{-1} \boldsymbol{\mathcal{C}}_t^{(i)} \boldsymbol{\Sigma}_{t|t-1}^{(i)}, \\
\ell_t^{(i)} &= -\frac{1}{2} \big[\ln |\boldsymbol{\Omega}_{t|t-1}^{(i)}| + (\widetilde{\mathcal{Y}}_t^{(i)})' (\boldsymbol{\Omega}_{t|t-1}^{(i)})^{-1} \widetilde{\mathcal{Y}}_t^{(i)} \big],
\end{aligned}$$

where $\boldsymbol{\mathcal{A}}_t^{(i)}$, $\boldsymbol{\mathcal{B}}_t^{(i)}$, $\boldsymbol{\mathcal{C}}_t^{(i)}$, and $\boldsymbol{\Omega}_{\mathcal{U}}^{(i)}$ are the state space matrices defined in Section S1 above, evaluated at values for the nonlinear states given by $\mathcal{V}_t^{(i)}$ and parameter values $\Psi_0^{(i)}$.[23]

S3.1.4 *Details of updating the sufficient statistics for the parameters* In $\mathcal{M}_0$, variances have $\mathcal{IG}$ conjugate priors and θ has a conjugate normal prior. This section argues conjugacy is preserved when the support of θ is on the unit circle.

Variance parameters Index the variance parameters by $\ell = \eta, \upsilon, \kappa, \zeta_\pi, \zeta_1, \zeta_2, \ldots, \zeta_{\mathcal{H}}$. As described in Section 3 of the paper, each variance parameter σ_ℓ^2 has a conjugate $\mathcal{IG}$ prior

[23]There are missing observations in the SPF inflation data the Kalman filter handles using standard methods. When observations are missing, the corresponding rows of $\boldsymbol{\mathcal{C}}_t^{(i)}$, $\mathcal{Y}_t$, and $\widetilde{\mathcal{Y}}_t^{(i)}$ are set to zero and pseudo-inverses and pseudo-determinants of $\boldsymbol{\Omega}_{t|t-1}^{(i)}$ are computed. For numerical stability, this is achieved using a Kalman filter in square root form as described in Lindsten, Bunch, Särkkä, Schön, and Godsill (2016) using fast-array methods developed by Kailath, Sayed, and Hassibi (2000).

that is characterized by sufficient statistics $\alpha_{t,\ell}^{(i)}$ and $\beta_{t,\ell}^{(i)}$, which represent the number of prior observations and the associated sum of squares. Updating the sufficient statistics at a given date t increases the number of observations by one, $\alpha_{t,\ell}^{(i)} = \alpha_{t-1,\ell}^{(i)} + 1$, and adds the most recent, squared observation to the scale parameter $\beta_{t,\ell}^{(i)} = \beta_{t-1,\ell}^{(i)} + (\Delta_{t,\ell}^{(i)})^2$, where $\Delta_{t,\ell}^{(i)}$ denotes the innovation of the corresponding ℓth parameter of particle i at date t. For $\ell = \eta$ or σ_η^2, $\Delta_{t,\eta}^{(i)} = \ln(\varsigma_{\eta,t}^{(i)})^2 - \ln(\varsigma_{\eta,t-1}^{(i)})^2$. Alternatively, in case of the variance of the irregular inflation gap, let $\Delta_{t,\zeta_\pi}^{(i)} = h_\pi(\mathcal{Y}_t - \mathcal{C}_t^{(i)}\mathcal{S}_t^{(i)})$, where h_π is a selection vector that isolates the first element of $\mathcal{Y}_t$, which is π_t, and similarly for the variances of the survey measurement errors.[24] Regarding the innovation variance of shocks to λ_t, $\ell = \kappa$, treating the $\mathcal{IG}$ prior as a conjugate prior holds only as an approximation because it neglects the rejection of shocks $\Delta_{t,\kappa} = \sigma_\kappa \kappa_t$ that push λ_t outside the zero-one interval.[25] The quality of this approximation depends, of course, on the frequency with which draws of $\Delta_{t,\kappa}$ yield λ_t outside its bounds, which is low for most of our particles.[26] Our Online Supplementary Appendix has robustness results to an alternative specification for the dynamics of λ_t that avoids the need for sampling from a limited support.

Constant AR1 inflation gap persistence parameter For a constant-θ, sufficient statistics of the $\mathcal{TN}$ prior are its mean, $m_{\theta,0}$, and variance, $V_{\theta,0}$. The prior is $\theta \sim \mathcal{TN}(m_{\theta,0}, V_{\theta,0} | \theta \in (-1,1))$, where $m_{\theta,0} = 0$ and $V_{\theta,0} = 1$. Suppose the support of θ was not truncated. In this case, the prior would be conjugate normal yielding a posterior that is also normal with mean and variance

$$m_{\theta,t}^{(i)} = \frac{m_{\theta,t-1}^{(i)} + V_{\theta,t-1}^{(i)} x_t^{(i)} y_t^{(i)}}{1 + V_{\theta,t-1}^{(i)} \left(x_t^{(i)}\right)^2}, \qquad V_{\theta,t}^{(i)} = \frac{V_{\theta,t-1}^{(i)}}{1 + V_{\theta,t-1}^{(i)} \left(x_t^{(i)}\right)^2}, \tag{A.34}$$

where

$$x_t^{(i)} = \frac{\varepsilon_{t-1}^{(i)}}{\varsigma_{v,t}^{(i)}} \quad \text{and} \quad y_t^{(i)} = \frac{\varepsilon_t^{(i)}}{\varsigma_{v,t}^{(i)}}, \tag{A.35}$$

given $\varepsilon_t^{(i)}$, $\varepsilon_{t-1}^{(i)}$, and $\varsigma_{v,t}^{(i)}$.[27] The recursion starts with $m_{\theta,0}^{(i)} = m_{\theta,0}$, and $V_{\theta,0}^{(i)} = V_{\theta,0}$.[28]

In our case of a *truncated* normal prior, conjugacy is preserved and sufficient statistics are updated using (A.34). Let the prior of the ith particle for θ at date t be given by

[24] Recall $\mathcal{S}_t^{(i)}$ is drawn from $\mathcal{N}(\mathcal{S}_t | \mathcal{S}_{t|t}^{(i)}, \boldsymbol{\Sigma}_{t|t}^{(i)})$ as in step 2.d) of the algorithm of Section S3.1.2.

[25] A similar caveat applies for $\mathcal{M}_2$ and $\mathcal{M}_3$ when θ_t is time-varying, but bounded inside the unit circle.

[26] We evaluate the probability of $\lambda_t \notin (0,1)$, given draws for λ_{t-1} from our smoother and full-sample estimate $\widehat{\sigma}_\kappa^2 = 0.008$ for $\mathcal{M}_2$. In $\mathcal{M}_0$, $\widehat{\sigma}_\kappa^2 = 0.005$ is smaller yielding lower probabilities of hitting the bounds. For 50% (90%) of all draws, the rejection probabilities < 4% (20%) with only a few exceptions.

[27] Draws of $\varepsilon_t^{(i)}$ and $\varepsilon_{t-1}^{(i)}$ are generated as described in step 2.d of Section S3.1.2.

[28] The updating formulas in (A.34), reflect the standard case of a Bayesian regression with normal priors and known residual variance as described, for example, by Hamilton (1994, Chapter 12).

$\mathcal{TN}(m^{(i)}_{\theta,t-1}, V^{(i)}_{\theta,t-1}|\theta \in (-1,1))$, its density function is

$$f(\theta|m^{(i)}_{\theta,t-1}, V^{(i)}_{\theta,t-1}) = \phi\left(\frac{\theta - m^{(i)}_{\theta,t-1}}{\sqrt{V^{(i)}_{\theta,t-1}}}\right) \cdot \frac{\mathbf{1}(-1<\theta<0)}{\Phi\left(\frac{1-m^{(i)}_{\theta,t-1}}{\sqrt{V^{(i)}_{\theta,t-1}}}\right) - \Phi\left(\frac{-1-m^{(i)}_{\theta,t-1}}{\sqrt{V^{(i)}_{\theta,t-1}}}\right)} \cdot \frac{1}{\sqrt{V^{(i)}_{\theta,t-1}}}$$

$$\propto \phi\left(\frac{\theta - m^{(i)}_{\theta,t-1}}{\sqrt{V^{(i)}_{\theta,t-1}}}\right) \cdot \mathbf{1}(-1<\theta<0),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal *p.d.f.* and *c.d.f.* and $\mathbf{1}(\cdot)$ is an indicator function. Since the likelihood of observing $y^{(i)}_t$ is $f(y^{(i)}_t|\theta, x^{(i)}_t) = \phi(y^{(i)}_t - \theta x^{(i)}_t)$, the posterior of θ retains the kernel of a truncated normal $f(\theta|y^{(i)}_t, x^{(i)}_t, m^{(i)}_{\theta,t-1}, V^{(i)}_{\theta,t-1}) \propto f(y^{(i)}_t|\theta, x^{(i)}_t) \cdot f(\theta|m^{(i)}_{\theta,t-1}, V^{(i)}_{\theta,t-1}) \propto \phi(\frac{\theta - m^{(i)}_{\theta,t}}{\sqrt{V^{(i)}_{\theta,t}}}) \cdot \mathbf{1}(-1<\theta<0)$, where $m^{(i)}_{\theta,t}$ and $V^{(i)}_{\theta,t}$ are given in (A.34).[29]

S3.2 *Particle filter for other model variants*

In $\mathcal{M}_0$, the AR1 inflation gap persistence parameter, θ, is constant and the SI weight λ_t is time-varying. Our particle learning filter algorithm is easily adapted to $\mathcal{M}_2$ and $\mathcal{M}_3$ that have θ_t. In these SSMs, θ_t is part of $\mathcal{V}_t$. Its innovation variance, σ^2_ϕ, is added to Ψ. Inference is analogous to the case of λ_t, as described above for $\mathcal{M}_0$.[30]

The particle learning filter when λ is a constant parameter When λ is a constant parameter, as in $\mathcal{M}_1$ and $\mathcal{M}_3$, it becomes part of Ψ. The particle learning algorithm estimates of λ restrict it to the $(0,1)$ interval. We chose a Beta prior, $\lambda \sim \text{Beta}(\alpha^\lambda_0, \beta^\lambda_0)$, which is conjugate and ensures $\lambda \in (0,1)$. With $\alpha^\lambda_0 = 1$ and $\beta^\lambda_0 = 1$, our prior matches a uniform distribution over the range of admissible values for λ.

In general, the Beta parameters α^λ and β^λ can be any positive real number. However, intuition about our updating procedure is made clear by limiting α^λ and β^λ to positive integers. In this case, the Beta distribution reflects Bayesian updating about the probability parameter of a Bernoulli experiment, denoted λ, after seeing a sequence of independent Bernoulli draws. Specifically, after observing $\mathtt{S}$ Bernoulli draws of success and $\mathtt{N}$ draws of failure, the Beta prior would be updated to $\lambda|\mathtt{S},\mathtt{N} \sim \text{Beta}(\alpha^\lambda_0 + \mathtt{S}, \beta^\lambda_0 + \mathtt{N})$ with posterior mean equal to $(\alpha^\lambda_0 + \mathtt{S})/(\alpha^\lambda_0 + \mathtt{S} + \beta^\lambda_0 + \mathtt{N})$.

We appeal to the Mankiw and Reis (2002) SI mechanism to update λ in the particle learning filter. Mankiw and Reis model λ to measure the fraction of forecasters that do not update. This is equivalent to observing a fraction of $\mathtt{S}/(\mathtt{S}+\mathtt{N})$ "successful" draws in a sequence of $\mathtt{S}+\mathtt{N}$ Bernoulli trials. At every date t, observations of a particle draw for $\lambda^{(i)}$, are treated as observing that a fraction $\lambda^{(i)}$ of forecasters had not updated its forecast, while a fraction $1-\lambda^{(i)}$ had. For particle i at time t with a prior $\text{Beta}(\alpha^{\lambda,(i)}_{t-1}, \beta^{\lambda,(i)}_{t-1})$,

[29] Given the definitions in (A.35), $f(\theta|\varepsilon^{(i)}_t, \varepsilon^{(i)}_{t-1}, \varsigma^{(i)}_{\upsilon,t}, m^{(i)}_{\theta,t-1}, V^{(i)}_{\theta,t-1}) = f(\theta|y^{(i)}_t, x^{(i)}_t, m^{(i)}_{\theta,t-1}, V^{(i)}_{\theta,t-1})$.

[30] The bounds on the support of $\theta_t \in (-1,1)$ is enforced using a procedure similar to λ_t in $\mathcal{M}_0$.

and given a parameter draw $\lambda^{(i)}$ at t, we update the sufficient statistics for the particle learning path of λ using $\alpha_t^{\lambda,(i)} = \alpha_{t-1}^{\lambda,(i)} + \lambda^{(i)}$ and $\beta_t^{\lambda,(i)} = \beta_{t-1}^{\lambda,(i)} + 1 - \lambda^{(i)}$.

In contrast to the above example of a sequence of Bernoulli draws, this update assigns the time t observation of $\lambda^{(i)}$ only with the weight of a single Bernoulli draw instead of $\mathtt{S} + \mathtt{N}$ draws. This causes the particle learning filter to take only limited signal at every date t. In principle, we could also scale up the relevance of each update by using $\alpha_t^{\lambda,(i)} = \alpha_{t-1}^{\lambda,(i)} + \lambda^{(i)} \cdot (\mathtt{S}+\mathtt{N})$ and $\beta_t^{\lambda,(i)} = \alpha_{t-1}^{\lambda,(i)} + (1 - \lambda^{(i)}) \cdot (\mathtt{S}+\mathtt{N})$ for some value of $\mathtt{S}+\mathtt{N}$, but this alternative had little effect on estimated values in practice.

S3.3 *A Rao–Blackwellized particle smoother*

This section describes how we utilize the particle smoothing methods developed by Lindsten et al. (2016), henceforth LBSSG, to generate "smoothed" estimates of linear and nonlinear states. Smoothed estimates reflect posteriors that condition on the full data sample, such as $p(\mathcal{S}_t|\mathcal{Y}^T)$ and $p(\mathcal{V}_t|\mathcal{Y}^T)$, while integrating out parameter uncertainty.

The goal is to simulate trajectories of $\mathcal{T}_t$ drawn from $p(\mathcal{T}^T|\mathcal{Y}^T)$. The first step factors the smoothing density $p(\mathcal{T}^T|\mathcal{Y}^T) = \int_\Psi p(\mathcal{T}^T|\mathcal{Y}^T, \Psi) p(\Psi|\mathcal{Y}^T)\, d\Psi$. There are well-known methods for simulating smoothed estimates of $\mathcal{T}_t$, given Ψ. Godsill, Doucet, and West (2004) proposed a forward-filtering-backward-simulation (FFBS) smoother to draw from $p(\mathcal{T}^T|\mathcal{Y}^T, \Psi)$ for a SSM that is not Rao–Blackwellized (and where $\mathcal{T}_t$ is not necessarily partitioned into $\mathcal{S}_t$ and $\mathcal{V}_t$). LBSSG adapt the FFBS particle smoother to the case when $\mathcal{T}_t$ is partitioned into $\mathcal{S}_t$ and $\mathcal{V}_t$, and $\mathcal{S}_t$ has been Rao–Blackwellized in the particle filter.

The FFBS smoother of LBSSG depends on a known Ψ. We wrap the methods of LBSSG into a problem of joint inference over $\mathcal{S}_t$, $\mathcal{V}_t$, and Ψ as follows: Parameter values are drawn from date T particle swarm of the particle learning filter. Conditional on each draw of Ψ, we apply the LBSSG smoother conditional to obtain draws of the smoothed states. In the final step, we integrate over the parameter draws:

$$p\big(\mathcal{V}^T, \mathcal{S}^T|\mathcal{Y}^T\big) = \int_\Psi p\big(\mathcal{V}^T, \mathcal{S}^T|\mathcal{Y}^T, \Psi\big)\, p\big(\Psi|\mathcal{Y}^T\big)\, d\Psi.$$

In sum, the FFBS particle smoothing algorithm consists of the following steps.

1. Run the particle learning filter, as described in Sections S3.1 and S3.2, to produce the full-sample particle learning swarm $\{\Psi^{(i)}, W_T^{(i)}\}_{i=1}^M$.
2. From $t = 1, \dots, T$, generate smoothed trajectories of $\{\widetilde{\mathcal{S}}_t^{(k)}, \widetilde{\mathcal{V}}_t^{(k)}\}_{k=1}^K$ by iterating
   (a) conditional on draws of $\widetilde{\Psi}^{(k)}$ from $\{\Psi^{(i)}, W_T^{(i)}\}_{i=1}^M$ with probability $W_T^{(k)}$, employ a Rao–Blackwellized auxiliary particle filter to generate the forward-filtering swarm $\{\mathcal{S}_{t|t}^{(n)}, \boldsymbol{\Sigma}_{t|t}^{(n)}, \mathcal{V}_{t|t}^{(n)}\}_{n=1}^N$ with weights $\{\widehat{W}_t^{(n)}\}_{n=1}^N$,
   (b) simulate $\{\widetilde{\mathcal{V}}_t^{(k)}\}_{t=1}^T$ backwards from $T-1$ by drawing from $p(\mathcal{V}^{t+1:T}|\mathcal{Y}^T, \Psi^{(k)})$, where $\mathcal{V}^{t+1:T}$ are the nonlinear states from date $t+1$ to T,

(c) and the final forward filtering step operates the Kalman filter forward to produce $\widetilde{\mathcal{S}}_t^{(k)}$ and $\widetilde{\boldsymbol{\Sigma}}_{t|t}^{(k)}$ conditional on $\widetilde{\mathcal{V}}_t^{(k)}$, $\mathcal{Y}^t$, and $\widetilde{\Psi}^{(k)}$.

In step 2(b), $p(\mathcal{V}^{t+1:T}|\mathcal{Y}^T, \Psi^{(k)})$ approximates the true density of smoothed $\mathcal{V}_t$, which is discussed in the paper and by LBSSG. Although draws from $p(\mathcal{V}^{t+1:T}|\mathcal{Y}^T, \Psi^{(k)})$ do not depend on $\widetilde{\mathcal{S}}_t^{(k)}$, it is an input into computing the probabilities of drawing $\widetilde{\mathcal{V}}^{t+1:T}$. Our results rest on running the particle learning filter on $M = 100{,}000$ particles and the FFBS smoother relies on $K = 1{,}000$ simulated trajectories, given $N = 10{,}000$ particles.

References

Anderson, B. D. O. and J. B. Moore (1979), *Optimal Filtering*. Prentice-Hall Information and System Sciences Series. Prentice-Hall, Englewood Cliffs, NJ. [9]

Anderson, E. W., E. R. McGrattan, L. P. Hansen, and T. J. Sargent (1996), "Mechanics of forming and estimating dynamic linear economies." In *Handbook of Computational Economics*, Vol. 1 (H. M. Amman, D. A. Kendrick, and J. Rust, eds.), Chapter 4, 171–252, Elsevier, Amsterdam. [9]

Carvalho, C. M., M. S. Johannes, H. F. Lopes, and N. G. Polson (2010), "Particle learning and smoothing." *Statistical Science*, 25 (1), 88–106. [14]

Cogley, T. and T. J. Sargent (2008), "Anticipated utility and RE as approximations of Bayesian decision making." *International Economic Review*, 49 (5), 185–221. [5]

Cogley, T. and A. M. Sbordone (2008), "Trend inflation, indexation, and inflation persistence in the new Keynesian Phillips curve." *American Economic Review*, 98 (5), 2101–2126. [5]

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007), "ABCs (and Ds) of understanding VARs." *American Economic Review*, 97 (3), 1021–1026. [7]

Godsill, S. J., A. Doucet, and M. West (2004), "Monte Carlo smoothing for nonlinear time series." *Journal of the American Statistical Association*, 99 (465), 156–168. [19]

Grassi, S. and T. Proietti (2010), "Has the volatility of U.S. inflation changed and how?" *Journal of Time Series Econometrics*, 2 (1), Article 6. [2]

Hamilton, J. D. (1994), *Time Series Analysis*. Princeton University Press, Princeton, NJ. [17]

Hansen, L. P. and T. J. Sargent (2007), *Robustness*. Princeton University Press, Princeton, NJ. [9]

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge. [9]

Justiniano, A. and G. E. Primiceri (2008), "The time-varying volatility of macroeconomic fluctuations." *American Economic Review*, 98 (3), 604–641. [5]

Kailath, T., A. H. Sayed, and B. Hassibi (2000), *Linear Estimation*. Prentice-Hall Information and System Sciences Series. Prentice-Hall, Upper Saddle River, NJ. [9, 16]

Komunjer, I. and S. Ng (2011), "Dynamic identification of dynamic stochastic general equilibrium models." *Econometrica*, 79 (6), 1995–2032. [6]

Kreps, D. (1998), "Anticipated utility and dynamic choice." In *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997* (E. K. Donal, P. Jacobs, M. I. Kamien, and N. L. Schwartz, eds.), 242–274, Cambridge University Press, Cambridge. [5]

Lindsten, F., P. Bunch, S. Särkkä, T. B. Schön, and S. J. Godsill (2016), "Rao–Blackwellized particle smoothers for conditionally linear Gaussian models." *IEEE Journal of Selected Topics in Signal Processing*, 10 (2), 353–365. [16, 19]

Lopes, H. F. and R. S. Tsay (2011), "Particle filters and Bayesian inference in financial econometrics." *Journal of Forecasting*, 30 (1), 168–209. [14]

Mankiw, N. G. and R. Reis (2002), "Sticky information versus sticky prices: A proposal to replace the new Keynesian Phillips curve." *Quarterly Journal of Economics*, 117 (4), 1295–1328. [18]

Muth, J. F. (1960), "Optimal properties of exponentially weighted forecasts." *Journal of the American Statistical Association*, 55 (290), 299–306. [2]

Pitt, M. K., R. dos Santos Silva, P. Giordani, and R. Kohn (2012), "On some properties of Markov chain Monte Carlo simulation methods based on the particle filter." *Journal of Econometrics*, 171 (2), 134–151. [16]

Pitt, M. K. and N. Shephard (1999), "Filtering via simulation: Auxiliary particle filters." *Journal of the American Statistical Association*, 94 (446), 590–599. [14]

Pitt, M. K. and N. Shephard (2001), "Auxiliary variable based particle filters." In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. de Freitas, and N. Gordon, eds.), Chapter 13, Springer, Berlin. [14]

Shephard, N. (2015), "Martingale unobserved component models." In *Unobserved Components and Time Series Econometrics* (S. J. Koopman and N. Shephard, eds.), 218–249, Oxford University Press, Oxford. [2]

Stock, J. H. and M. W. Watson (2007), "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39 (S1), 3–33. [2, 12]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 18 September, 2017; final version accepted 8 May, 2020; available online 2 June, 2020.