Online Appendix of the paper Walk on the Wild Side: Temporarily Unstable Paths and Multiplicative Sunspots

By Guido Ascari and Paolo Bonomolo and Hedibert F. Lopes *

I. Implementation: The general solution

As in LS, we follow the approach of Sims (2002) and we write a general linear RE system as:

$$\Gamma_0 y_t = \Gamma_1 y_{t-1} + \Psi \varepsilon_t + \Pi \eta_t, \tag{A1}$$

where y_t is the vector of the n endogenous variables (including the expectations as in (2)), ε_t is the vector of the h exogenous fundamental shocks, and η_t is the vector of the $k \leq n$ RE forecast errors. For simplicity, we assume that Γ_0 is invertible,[1] so to write as in (12) :

$$y_t = \Gamma_1^* y_{t-1} + \Psi^* \varepsilon_t + \Pi^* \eta_t. \tag{A2}$$

The multivariate case is a relatively straightforward extension of the simple case, so the description follows similar steps as above, involving: (i) parameterizing the system using M (now a matrix); (ii) introducing time variation in M, and (iii) imposing stability. As usual, however, first we need to decouple the system through a variable transformation.

Partitioning. As in the main text, use Jordan decomposition to partition the system, and define the vector of transformed variables $\tilde{y}_t = J^{-1} y_t$. Let the ith element of $\tilde{y}_t$ be $\tilde{y}_{it}$, the ith element on the principal diagonal of Λ be λ_i and denote the ith row of $J^{-1}\Pi^*$ and $J^{-1}\Psi^*$ by $\left[J^{-1}\Pi^*\right]_i$ and $\left[J^{-1}\Psi^*\right]_i$, respectively. The model can then be written as a collection of AR(1) processes as in the univariate case: $\tilde{y}_{it} = \lambda_i \tilde{y}_{it-1} + \left[J^{-1}\Psi^*\right]_i \varepsilon_t + \left[J^{-1}\Pi^*\right]_i \eta_t$. Order the eigenvalues (and the corresponding eigenvectors) in descending order, and partition the system into two blocks of dimensions $(n-k)$ and k, respectively. As explained in the main text, we depart from Sims (2002) and LS because we partition the system as in (13), according to the number of forward-looking variables/expectation errors, rather than the number of explosive eigenvalues. Let m be the number of explosive eigenvalues (i.e., such that $\lambda_i \geq 1$). As usual, we assume that the number of explosive eigenvalues is smaller or equal to the number

* Ascari: University of Oxford, University of Pavia and Bank of Finland.Address: Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK. Email: guido.ascari@economics.ox.ac.uk. Bonomolo: De Nederlandsche Bank. Address: De Nederlandsche Bank, Economics and Research Division, Postbus 98, 1000 AB Amsterdam, The Netherlands. Email: p.bonomolo@dnb.nl. Lopes: INSPER. Address: INSPER - Institute of Education and Research, Quatá Street, 300 - Vila Olímpia, São Paulo - SP, 04546-042, Brasil. Email: hedibertfl@insper.edu.

[1] This is the case in the LS's model, which we will use in our empirical analysis below. If Γ_0 is singular, it is trivial to generalize the method to use the Schur decomposition (QZ).

of forecast errors, to rule out instability. Hence, the first $(n-k)$ rows only contain stable eigenvalues, while the last k rows contain both $(k-m)$ stable and m unstable eigenvalues. Hence, we do not need to impose any stability condition on the first block of the system (13), but we do need to do so on the second block of the system, i.e., (14).

Parameterization. Note that the system is decoupled, so it is just a collection of independent AR(1) processes. Each row in (14) corresponds to our simple example above (2). As for the case of the simple model, it is possible to parameterize the fundamental solutions, i.e., where the expectation error is just a function of the structural shock, by modifying the stability condition under determinacy. In matrix notation, the usual stability condition under determinacy would be $J_{\mu 2}\left[\Psi^* \varepsilon_t + \Pi^* \eta_t\right] = 0$, and, as in the simple case, we modify it to $(I+M)J_{\mu 2}\Psi^*\varepsilon_t = -J_{\mu 2}\Pi^*\eta_t$, when we restrict the matrix M to be diagonal, with M_i being the ith element on the principal diagonal of M. Hence:

$$\tilde{y}_{k,t} = \Lambda_2 \tilde{y}_{k,t-1} + J_{\mu 2}\Psi^*\varepsilon_t - (I+M)J_{\mu 2}\Psi^*\varepsilon_t = \Lambda_2 \tilde{y}_{k,t-1} - MJ_{\mu 2}\Psi^*\varepsilon_t. \tag{A3}$$

Iterate (A3) backward to find:

$$\tilde{y}_{k,t} = -M \sum_{i=0}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i}. \tag{A4}$$

This expression corresponds to (4), and like (4), it exists assuming that we start from steady state (there exists a time 0, such that $\tilde{y}_{-k,i} = \varepsilon_{-i} = \eta_{-i} = 0, \forall i \geq 0$). Moreover, some of the solutions for $\tilde{y}_{i,k,t}$ in (A4) will be stable and some will be unstable, depending on the values of the M_i's and on the stability properties of the system, i.e., depending on the values of the $\lambda_{2,i}$'s, where $\lambda_{2,i}$ is the ith element on the principal diagonal of Λ_2.

Time variation. Assume now that the M_i elements on the principal diagonal of the matrix M are changing over time following independently distributed and uncorrelated stochastic processes. Our proposed solution is then:

$$\tilde{y}_{k,t} = -M_t \sum_{i=0}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i}, \tag{A5}$$

which corresponds to (6). Note that in each period t, the solution just depends on the current realization of M_t. A solution pins down the expectations errors, actually $J_{\mu 2}\Pi^*\eta_t$. As in Sims (2002), a solution pins down the expectations errors, actually $J_{\mu 2}\Pi^*\eta_t$. Plugging (A5) into (14) yields:

$$J_{\mu 2}\Pi^*\eta_t = -(I+M_t)J_{\mu 2}\Psi^*\varepsilon_t - (M_t - M_{t-1}) \sum_{i=1}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i}. \tag{A6}$$

The RE condition implies $E_{t-1}\left(J_{\mu 2}\Pi^*\eta_t\right) = 0$, so that each $M_{i,t}$ must be: 1) a martingale; and 2) uncorrelated with ε_t. Once more, it is easy to recognize two particular solutions: 1) the forward-looking solution, given by $M_t = 0 => \tilde{y}^F_{k,t} = 0 => \eta_t = -\left(J_{\mu 2}\Pi^*\right)^{-1} J_{\mu 2}\Psi^*\varepsilon_t, \forall t$; and 2) the backward-looking solution, given by $M_t = -I => \tilde{y}^B_{k,t} = \sum_{i=0}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i}$ and $\eta_t = 0, \forall t$. The forward-looking solution always exists and it is always (under our assumption) a stable solution: it is the only stable one under determinacy ($m = k$), while it is one out of many possible stable ones under indeterminacy ($m < k$). However, in this latter case, the forward-looking solution is a special one given how we partition the system: it coincides with the minimum state variable solution, because it delivers a solution which is just a linear function of the state variables.

Then the solution to the system of disconnected difference equations (A3) can be written recursively almost as in Blanchard (1979), but actually using only the backward-looking variable $\tilde{y}^B_{k,t}$ as:

$$\tilde{y}_{k,t} = -M_t \sum_{i=0}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i} = -M_t \tilde{y}^B_{k,t} \tag{A7}$$

so that:

$$\tilde{y}^B_{k,t} = \Lambda_2 \tilde{y}^B_{k,t-1} + J_{\mu 2}\Psi^*\varepsilon_t \tag{A8}$$

$$\tilde{y}_{k,t} = -M_t \tilde{y}^B_{k,t} = -M_t\left(\Lambda_2 \tilde{y}^B_{k,t-1} + J_{\mu 2}\Psi^*\varepsilon_t\right) \tag{A9}$$

which are (15) and (16) in the main text.

Note that since: $\tilde{y}^B_{k,t} = \sum_{i=0}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i} = J^{-1}\Psi^*\varepsilon_t + \sum_{i=1}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i}$, the expectation error could be written as:

$$\begin{aligned} J_{\mu 2}\Pi^*\eta_t &= -(I + M_t)J_{\mu 2}\Psi^*\varepsilon_t - (M_t - M_{t-1})\sum_{i=1}^{t-1} \Lambda_2^i \left(J_{\mu 2}\Psi^*\right)\varepsilon_{t-i} \\ &= -(I + M_t)J_{\mu 2}\Psi^*\varepsilon_t - (M_t - M_{t-1})\left(\tilde{y}^B_{k,t} - J_{\mu 2}\Psi^*\varepsilon_t\right) \\ &= -(I + M_{t-1})J_{\mu 2}\Psi^*\varepsilon_t - (M_t - M_{t-1})\,\tilde{y}^B_{k,t} \end{aligned}$$

which yields (17), assuming that the (kxk) matrix $J_{\mu 2}\Pi^*$ is invertible.

We discuss stability in the main text. Again, as in the simple model, and we impose stability by allowing only particular processes for $M_{i,t}$'s.

Recompose the system and solve for original variables. Having solved for the forward-looking variables, we now need to recompose the system from the original partition. First, we need to substitute for $J_{\mu 1}\left[\Psi^*\varepsilon_t + \Pi^*\eta_t\right]$ into (13), given the η_t implied by our proposed solution from (17). Substitute (A9) in the system (13), adding the auxiliary variable

$\tilde{y}^B_{k,t}$:

$$\begin{bmatrix} \tilde{y}_{(n-k),t} \\ ((n-k)\times 1) \\ \tilde{y}_{k,t} \\ (k\times 1) \\ \tilde{y}^B_{k,t} \\ (k\times 1) \end{bmatrix} = \begin{bmatrix} \Lambda_1 & \mathbf{0} & \mathbf{0} \\ ((n-k)\times(n-k)) & ((n-k)\times k) & ((n-k)\times k) \\ \mathbf{0} & \mathbf{0} & -M_t\Lambda_2 \\ (k\times(n-k)) & (k\times k) & (k\times k) \\ \mathbf{0} & \mathbf{0} & \Lambda_2 \\ (k\times(n-k)) & (k\times k) & (k\times k) \end{bmatrix} \begin{bmatrix} \tilde{y}_{(n-k),t-1} \\ ((n-k)\times 1) \\ \tilde{y}_{k,t-1} \\ (k\times 1) \\ \tilde{y}^B_{k,t-1} \\ (k\times 1) \end{bmatrix} + \\ + \begin{bmatrix} J_{\mu 1}\left[\Psi^*\varepsilon_t + \Pi^*\eta_t\right] \\ ((n-k)\times 1) \\ -M_t J_{\mu 2}\Psi^*\varepsilon_t \\ (k\times 1) \\ J_{\mu 2}\Psi^*\varepsilon_t \\ (k\times 1) \end{bmatrix}$$

Then the problem is to pin down $J_{\mu 1}\Pi^*\eta_t$, but we know η_t, given our proposed solution from (17), so:

$$\begin{aligned} & J_{\mu 1}\left[\Psi^*\varepsilon_t + \Pi^*\eta_t\right] \\ = \ & J_{\mu 1}\left[\Psi^*\varepsilon_t + \Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}\left[-(I+M_{t-1})J_{\mu 2}\Psi^*\varepsilon_t - \left(M_t - M_{t-1}\right)\tilde{y}^B_{k,t}\right]\right] \\ = \ & J_{\mu 1}\left[\Psi^* - \Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}(I+M_{t-1})J_{\mu 2}\Psi^*\right]\varepsilon_t - J_{\mu 1}\Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}\left(M_t - M_{t-1}\right)\tilde{y}^B_{k,t} \end{aligned}$$

Then given (A8), we can write:

$$\begin{aligned} & J_{\mu 1}\left[\Psi^*\varepsilon_t + \Pi^*\eta_t\right] \\ = \ & J_{\mu 1}\left[\Psi^* - \Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}(I+M_{t-1})J_{\mu 2}\Psi^*\right]\varepsilon_t + \\ - \ & J_{\mu 1}\Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}\left(M_t - M_{t-1}\right)\left(\Lambda_2\tilde{y}^B_{k,t-1} + J_{\mu 2}\Psi^*\varepsilon_t\right) \\ = \ & J_{\mu 1}\left[\Psi^* - \Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}(I-M_t)J_{\mu 2}\Psi^*\right]\varepsilon_t - J_{\mu 1}\Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}\left(M_t - M_{t-1}\right)\Lambda_2\tilde{y}^B_{k,t-1} \end{aligned}$$

So we can write:

$$J_{\mu 1}\left[\Psi^*\varepsilon_t + \Pi^*\eta_t\right] = A_t\varepsilon_t - B_{t,t-1}\tilde{y}^B_{k,t-1}, \tag{A10}$$

where A_t is the $(n-k)\times l$ matrix and $B_{t,t-1}$ is a $(n-k)\times k$ matrix, respectively given by (20) and (21) in the main text, that is:

$$A_t = J_{\mu 1}\left[\Psi^* - \Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}(I+M_t)J_{\mu 2}\Psi^*\right]; \tag{A11}$$

$$B_{t,t-1} = J_{\mu 1}\Pi^*\left(J_{\mu 2}\Pi^*\right)^{-1}\left(M_t - M_{t-1}\right)\Lambda_2 \tag{A12}$$

The final system is therefore:

$$\begin{aligned}\tilde{y}_{(n-k),t} &= \Lambda_1 \tilde{y}_{(n-k),t-1} - B_{t,t-1}\tilde{y}^B_{k,t-1} + A_t\varepsilon_t \\ \tilde{y}_{k,t} &= -M_t\Lambda_2\tilde{y}^B_{k,t-1} - M_t J_{\mu 2}\Psi^*\varepsilon_t \\ \tilde{y}^B_{(n-k),t} &= \Lambda_1\tilde{y}^B_{(n-k),t-1} + J_{\mu 1}\Psi^*\varepsilon_t \\ \tilde{y}^B_{k,t} &= \Lambda_2\tilde{y}^B_{k,t-1} + J_{\mu 2}\Psi^*\varepsilon_t,\end{aligned}$$

which in matrix notation is:

$$\text{(A13)} \qquad \begin{bmatrix} \tilde{y}_{(n-k),t} \\ \tilde{y}_{k,t} \\ \tilde{y}^B_{(n-k),t} \\ \tilde{y}^B_{k,t} \end{bmatrix} = \underbrace{\begin{bmatrix} \Lambda_1 & \mathbf{0} & \mathbf{0} & -B_{t,t-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -M_t\Lambda_2 \\ \mathbf{0} & \mathbf{0} & \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Lambda_2 \end{bmatrix}}_{G^*} \begin{bmatrix} \tilde{y}_{(n-k),t-1} \\ \tilde{y}_{k,t-1} \\ \tilde{y}^B_{(n-k),t-1} \\ \tilde{y}^B_{k,t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} A_t \\ -M_t J_{\mu 2}\Psi^* \\ J_{\mu 1}\Psi^* \\ J_{\mu 2}\Psi^* \end{bmatrix}}_{H^*} \varepsilon_t.$$

Finally, to recover the original variables, use $= \tilde{y}_t = J^{-1}y_t$ to obtain (18) in the main text.

II. The econometric strategy

Regarding the structural parameters of the model, collected in the vector θ, as well as the latent states, the inference is fully Bayesian. The time-varying characteristic of the latent state M_t leads to a non-linear and analytically intractable non-Gaussian likelihood function for the unknowns. This motivates the use of the Sequential Monte Carlo strategy described below.

A. Preliminaries

The class of solutions we propose in equation (18), parametrized by the matrix M_t, has state space representation (24) that we repeat below for convenience:

$$\text{(A14)} \qquad \begin{cases} D_t = c + Fl_t + v_t & v_t \sim N(\mathbf{0},\Sigma_v) \\ l_t = G_t l_{t-1} + H_t\varepsilon_t & \varepsilon_t \sim N(\mathbf{0},\Sigma_\varepsilon) \end{cases}$$

D_t is the vector with data at time t, and $D_{m:n}$ is the set of observations from m to n for $m \leq n$. The parameters of the model are collected in the vector $\theta = (\theta_1, \theta_2)$, where we group in θ_1 all the parameters other than the variances and the covariances of the shocks, which are in turn collected in the vector θ_2. Finally, we assume that the dynamics of M_t are described by a transition law:

$$\text{(A15)} \qquad M_t = f(M_{t-1}, \zeta_t)$$

where ζ_t is a multiplicative sunspot shock. The properties of the stochastic process for M_t are discussed in the paper.

Our econometric strategy is based on sequential learning: suppose the posterior distribution of the unknowns is approximated at time $t-1$ by a set of particles $\{(l_{t-1}, M_{t-1}, \theta_1, \theta_2)^{(i)}\}_{i=1}^N$ and associated weights $\{w_{t-1}^{(i)}\}_{i=1}^N$. Given the new observed data D_t, we want to generate an updated set of particles $\{(l_t, M_t, \theta_1, \theta_2)^{(i)}\}_{i=1}^N$ and weights $\{w_t^{(i)}\}_{i=1}^N$ that approximate the posterior distribution:

$$p\left(l_t, M_t, \theta_1, \theta_2 | D_{1:t}\right). \tag{A16}$$

The way we group the latent processes (distinguishing M_t from all other states l_t) and the parameters (dividing them in θ_1 and θ_2) has a specific reason: as a general principle of our econometric strategy, we implement analytical computation whenever it is possible. To this aim, note that given a value for M_t, the state space (A14) is linear and Gaussian: we can compute the posterior distribution of the latent processes in l_t analytically, using the Kalman filter. Moreover, an analytical expression for the posterior distribution can also be derived for some of the parameters that we collect in θ_2. For DSGE models, this is typically the case for the variances and covariances of the shocks, when the prior distributions are Inverse Gamma or Inverse Wishart. Then, following Carvalho et al. (2010), we keep track of a set of sufficient statistics collected in s_t that we will use to update the posterior distribution of θ_2.

To approximate the posterior distribution of the parameters in θ_1, we use the Liu and West (2001) filter. Since this method uses mixtures of Normal distributions we make sure that all the parameters have the right support, that is from $-\infty$ to $+\infty$. Then, we define a new vector ϕ where each element of θ_1 is appropriately transformed when needed. In the description of the algorithm, we will add a time t subscript to this parameter, writing ϕ_t. This notation is introduced simply to reinforce the notion that sequential inference regarding ϕ is performed at time t, and it does not mean that the parameters are time-varying.

B. The particle filter

The algorithm we use is based on two main steps: an *updating* step, in which an appropriate number of particles N is drawn from an importance distribution $q\left(\vartheta_t, M_t, \theta_1, \theta_2 | D_{1:t}\right)$, and a *re-weighting* step in which the weights are computed as:

$$w_t^{(i)} = \frac{p\left(l_t^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \theta_2^{(i)} | D_{1:t}\right)}{q\left(l_t^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \theta_2^{(i)} | D_{1:t}\right)}. \tag{A17}$$

Step 1: Drawing from the importance distribution

Drawing from the importance distribution involves two sub-steps, following the schema in Pitt and Shephard (1999): a resampling step in which we select "the most fit particles", and the actual propagation step in which these particles are updated.

Resampling. Once new data have arrived, we start selecting the particles with higher predictive ability. We perform a resampling step using weights $\tilde{w}_t^{(i)}$ proportional to:

$$\tilde{w}_t^{(i)} \propto w_{t-1}^{(i)} p\left(D_t | l_{t-1}^{(i)}, g_M(M_{t-1}^{(i)}), m_{t-1}^{(i)}, \theta_2^{(i)}\right) \tag{A18}$$

Following Pitt and Shephard (1999) and Liu and West (2001), the predictive likelihood in equation (A18) is conditional on $g_M(M_{t-1}^{(i)})$, that is a *best guess* of M_t^i at time $t-1$ like $E\left(M_t|M_{t-1}^{(i)}\right)$, and on m_{t-1}^i defined as:[2]

$$m_{t-1}^{(i)} = a\phi_{t-1}^{(i)} + (1-a)\bar{\phi}_{t-1} \tag{A19}$$

where $\bar{\phi}_{t-1}$ is the weighted sample mean of $\phi_{t-1}^{(i)}$. Define also V_{t-1} as the sample weighted covariance matrix of $\phi_{t-1}^{(i)}$, that we will use later. Then, given the state space (A14), the predictive likelihood is a Normal distribution with mean $\hat{f}_t^{(i)}$ and variance $\hat{Q}_t^{(i)}$ where:

$$\hat{f}_t^{(i)} = \hat{c}^{(i)} + \hat{F}\hat{G}_{t-1}^{(i)} l_{t-1}^{(i)} \tag{A20}$$

$$\hat{Q}_t^{(i)} = \hat{F}\left(\hat{G}_{t-1}^{(i)} C_{t-1}^{(i)} \hat{G}_{t-1}^{(i)\prime} + \hat{H}_{t-1}^{(i)} \Sigma_\varepsilon^{(i)} \hat{H}_{t-1}^{(i)\prime}\right)\hat{F}' \tag{A21}$$

and $C_{t-1}^{(i)}$ is the variance of the latent process $l_{t-1}^{(i)}$. Note that the matrices $\hat{F}$, $\hat{G}_{t-1}^{(i)}$ and $\hat{H}_{t-1}^{(i)}$ and the vector $\hat{c}^{(i)}$ are functions of $g_M(M_{t-1}^{(i)})$ and of the parameters in $m_{t-1}^{(i)}$.

At this point, we have a set of resampled particles that, for convenience, we accentuate with a tilde: $\{(\tilde{l}_{t-1}, \tilde{M}_{t-1}, \tilde{m}_{t-1}, \tilde{\theta}_2, \tilde{s}_{t-1}, \tilde{C}_{t-1}, \tilde{f}_t, \tilde{Q}_t)^{(i)}\}_{i=1}^N$.

Propagation. The resampled particles are then propagated starting from the set of parameters $\phi_t^{(i)}$. Following the schema of Liu and West (2001), we update this vector drawing its new values from the normal distribution:

$$\phi_t^{(i)} \sim N\left(\tilde{m}_{t-1}^{(i)}, (1-a^2)V_{t-1}\right). \tag{A22}$$

Then, we proceed with the propagation of $M_{1,t}^{(i)}$ from the distribution implied by its law of motion (A15):

$$M_t^{(i)} \sim p\left(M_t|\tilde{M}_{t-1}^{(i)}, \phi_t^{(i)}, \tilde{\theta}_2^{(i)}\right) \tag{A23}$$

Given $M_t^{(i)}$ the state space (A14) becomes linear and Gaussian. We can draw $l_t^{(i)}$ from its

[2]The parameter a in equation (A19), which accounts for the amount of shrinkage, is suggested to be set between 0.974 and 0.995 (see Liu and West, 2001, for details)

posterior distribution:

$$l_t^{(i)} \sim p\left(l_t|\tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \phi_t^{(i)}, \tilde{\theta}_2^{(i)}, D_t\right) \tag{A24}$$

that is a Normal distribution with mean $\mu_t^{(i)}$ and variance $C_t^{(i)}$ computed through the Kalman filter recursion:

$$f_t^{(i)} = c^{(i)} + FG_t^{(i)}\tilde{l}_{t-1}^{(i)} \tag{A25}$$

$$Q_t^{(i)} = F\left(G_t^{(i)}\tilde{C}_{t-1}^{(i)}G_t^{(i)\prime} + H_t^{(i)}\tilde{\Sigma}_\varepsilon^{(i)}H_t^{(i)\prime}\right)F' \tag{A26}$$

$$\mu_t^{(i)} = G_t^{(i)}\tilde{l}_{t-1}^{(i)} + \left(G_t^{(i)}\tilde{C}_{t-1}^{(i)}G_t^{(i)\prime} + H_t^{(i)}\tilde{\Sigma}_\varepsilon^{(i)}H_t^{(i)\prime}\right)F'\left(Q_t^{(i)}\right)^{-1}\left(D_t - f_t^{(i)}\right) \tag{A27}$$

$$\begin{aligned} C_t^{(i)} &= \left(G_t^{(i)}\tilde{C}_{t-1}^{(i)}G_t^{(i)\prime} + H_t^{(i)}\tilde{\Sigma}_\varepsilon^{(i)}H_t^{(i)\prime}\right) + \\ &- \left(G_t^{(i)}\tilde{C}_{t-1}^{(i)}G_t^{(i)\prime} + H_t^{(i)}\tilde{\Sigma}_\varepsilon^{(i)}H_t^{(i)\prime}\right)F'\left(Q_t^{(i)}\right)^{-1}F\left(G_t^{(i)}\tilde{C}_{t-1}^{(i)}G_t^{(i)\prime} + H_t^{(i)}\tilde{\Sigma}_\varepsilon^{(i)}H_t^{(i)\prime}\right)\end{aligned} \tag{A28}$$

Note that the matrices F, $G_t^{(i)}$ and $H_t^{(i)}$, and the vector $c^{(i)}$ are functions of $M_t^{(i)}$ and of the updated parameters $\phi_t^{(i)}$. Then, the mean and the covariance matrix of the predictive distribution, respectively $f_t^{(i)}$ and $Q_t^{(i)}$, are different from those defined in (A20) and (A21).

Finally, we propagate the vector $\theta_2^{(i)}$ following the Particle Learning approach of Carvalho et al. (2010). The latent processes $l_t^{(i)}$ and $M_t^{(i)}$ and the parameters $\phi_t^{(i)}$ are used to update the set of sufficient statistics $s_t^{(i)}$.[3] Hence, we can draw $\theta_2^{(i)}$ from its posterior distribution:

$$\theta_2^{(i)} \sim p\left(\theta_2|s_t^{(i)}\right) \tag{A29}$$

We have drawn a new set of particles from the importance distribution obtained combining equations (A18), (A22), (A23), (A24) and (A29).

Step 2: Re-weighting the particles

In order to approximate the target density, we need to compute the appropriate weight for each particle, according to equation (A17).

Start from the joint posterior distribution (A16) which is proportional to:

$$p\left(l_t, M_t, \theta|D_{1:t}\right) \propto p\left(D_t|l_t, M_t, \theta\right)p\left(l_t, M_t, \theta|D_{1:(t-1)}\right), \tag{A30}$$

[3] For example, if the variance of a shock is a priori distributed as an Inverse Gamma, to compute the conjugate posterior we need the sum of the squared errors.

where the second term on the right-hand side is written as

$$\begin{aligned} & p\left(l_t, M_t, \theta | D_{1:(t-1)}\right) = \\ = & \int p\left(l_t, M_t, \theta | l_{1:(t-1)}, M_{1:(t-1)}\right) p\left(l_{1:(t-1)}, M_{1:(t-1)} | D_{1:(t-1)}\right) dl_{1:(t-1)} dM_{1:(t-1)} \\ \approx & \sum_{i=1}^{N} w_{t-1}^{(i)} p\left(l_t, M_t, \theta | l_{1:(t-1)}^{(i)}, M_{1:(t-1)}^{(i)}\right). \end{aligned} \tag{A31}$$

Consequently, the posterior is approximated by

$$p\left(l_t, M_t, \theta | D_{1:t}\right) \propto \sum_{i=1}^{N} w_{t-1}^{(i)} p\left(D_t | l_t, M_t, \theta\right) p\left(l_t, M_t, \theta | l_{1:(t-1)}^{(i)}, M_{1:(t-1)}^{(i)}\right). \tag{A32}$$

Assuming that the latent processes are Markov chains, we can write the numerator in equation (A17) as:

$$p\left(l_t^{(i)}, M_t^{(i)}, \theta^{(i)} | D_{1:t}\right) = w_{t-1}^{(i)} p\left(D_t | l_t^{(i)}, M_t^{(i)}, \theta^{(i)}\right) p\left(l_t^{(i)}, M_t^{(i)}, \theta^{(i)} | l_{t-1}^{(i)}, M_{t-1}^{(i)}\right). \tag{A33}$$

Following Carvalho et al. (2010), we compute the weights before propagating the parameters in θ_2. Taking this into account and combining equations (A18), (A22), (A23), (A24), (A29) and (A33) in equation (A17), we get:[4]

$$w_t^{(i)} \propto \frac{p\left(D_t | l_t^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}\right) p\left(l_t^{(i)} | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}\right)}{p\left(D_t | \tilde{l}_{t-1}^{(i)}, g_M(\tilde{M}_{t-1}^{(i)}), \tilde{m}_{t-1}^{(i)}, \tilde{\theta}_2^{(i)}\right) p\left(l_t^{(i)} | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}, D_t\right)} \tag{A34}$$

Note that the density $p\left(l_t^{(i)} | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}, D_t\right)$ in the denominator can be rewritten as

$$p\left(l_t^{(i)} | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}, D_t\right) = \frac{p\left(D_t | l_t^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}\right) p\left(l_t^{(i)} | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}\right)}{p\left(D_t | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}\right)} \tag{A35}$$

Substituting this equation in (A34), we find that the weights to approximate the joint posterior distribution at time t are:

$$w_t^{(i)} \propto \frac{p\left(D_t | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)}\right)}{p\left(D_t | \tilde{l}_{t-1}^{(i)}, g_M(\tilde{M}_{t-1}^{(i)}), \tilde{m}_{t-1}^{(i)}, \tilde{\theta}_2^{(i)}\right)}. \tag{A36}$$

[4]The weights are expressed as "proportional to" instead of "equal to" because they need to be normalized such that their sum is equal to one.

At the numerator, we have the Normal distribution with mean $f_t^{(i)}$ and covariance matrix $Q_t^{(i)}$ defined in equations (A25) and (A26). The distribution at the denominator is the Normal with mean $\tilde{f}_t^{(i)}$ and covariance matrix $\tilde{Q}_t^{(i)}$ defined in (A20) and (A21), and resampled according to weights $\tilde{w}_t^{(i)}$ computed in (A18). Both densities are evaluated in D_t.

Equation (A36) is very intuitive: the weight of each particle is computed comparing two predictive likelihoods. The particle i has a higher weight if, after propagation of $M_t^{(i)}$ and $\theta_1^{(i)}$, this leads to a higher improvement in predicting D_t.

Step 3 (optional): Resampling

The approximation of the posterior distribution obtained in the two steps described above is good if the the particle weights in (A36) are Uniformly distributed. It is well known in the literature that the variance of the distribution of the weights tends to increase over time since a subset of particles will have higher predictive power. Then, an additional resampling step using the weights computed in (A36) can be added to mitigate this problem. After a resampling step has been performed, all weights are set equal to $1/N$.

Usually the final resampling step is implemented when a certain criterion suggests that the distribution of weights became too uneven. A common practice is to check the *effective sample size* defined as:

$$N_t^e = \left(\sum_{i=1}^{N} \left(w_t^{(i)} \right)^2 \right)^{-1} . \tag{A37}$$

N_t^e takes values from 1 (very uneven distribution) to N (Uniform distribution), so the resampling step is performed when N_t^e is less than a certain threshold $\bar{N}$.

The procedure to implement our particle filter is summarized in the algorithm below.

THE ALGORITHM

Initialization: t=0

Draw a set of particles $\{(l_0, M_0, \theta_1, \theta_2, s_0, C_0)^{(i)}\}_{i=1}^N$ from a prior

Recursion: for $t = 1, 2, ...T$ repeat steps 1 to 6

1. Approximate $p(\phi|D_{0:(t-1)})$
   1a) Consider a transformation of the vector θ_1 and call it ϕ_t
   1b) Compute the weighted sample mean $\bar{\phi}_{t-1}$ and the covariance matrix V_{t-1}
   1c) Compute $m_{t-1}^{(i)} = a\phi_{t-1}^{(i)} + (1-a)\bar{\phi}_{t-1}$
2. Resample
   2a) Compute $\tilde{w}_t^{(i)} \propto w_{t-1}^{(i)} p\left(D_t | l_{t-1}^{(i)}, g_M(M_{t-1}^{(i)}), m_{t-1}^{(i)}, \theta_2^{(i)}\right)$
   2b) Resample $\{(l_{t-1}, M_{t-1}, m_{t-1}, \theta_2, s_t, C_t)^{(i)}\}_{i=1}^N$ with weights $\tilde{w}_t^{(i)}$
   Let the new particles be $\{(\tilde{l}_{t-1}, \tilde{M}_{t-1}, \tilde{m}_{t-1}, \tilde{\theta}_2, \tilde{s}_{t-1}, \tilde{C}_{t-1})^{(i)}\}_{i=1}^N$.
3. Propagate
   3a) Sample $\phi_t^{(i)}$ from $N\left(\tilde{m}_{t-1}^{(i)}, (1-a^2)V_{t-1}\right)$
   3b) Sample $M_t^{(i)}$ from $p\left(M_t | \tilde{M}_{t-1}^{(i)}, \phi_t^{(i)}, \tilde{\theta}_2^{(i)}\right)$
   3c) Sample $l_t^{(i)}$ from $N\left(\mu_t^{(i)}, C_t^{(i)}\right)$
   where $\mu_t^{(i)}$ and $C_t^{(i)}$ are defined in (A27) and (A28).
4. Compute new weights

$$w_t^{(i)} \propto \frac{p(D_t | \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \theta_1^{(i)}, \tilde{\theta}_2^{(i)})}{p(D_t | \tilde{l}_{t-1}^{(i)}, g_M(\tilde{M}_{t-1}^{(i)}), \tilde{m}_{t-1}^{(i)}, \tilde{\theta}_2^{(i)})}.$$

5. Update sufficient statistics and propagate θ_2
   5a) Compute $s_t^{(i)} = \mathcal{S}\left(l_t^{(i)}, \tilde{l}_{t-1}^{(i)}, M_t^{(i)}, \tilde{M}_t^{(i)}, \phi_t^{(i)}, D_t\right)$
   5b) Sample $\theta_2^{(i)}$ from $p\left(\theta_2 | s_t^{(i)}\right)$
6. Decide to resample or not
   if $\bar{N} < \left(\sum_{i=1}^N \left(w_t^{(i)}\right)^2\right)^{-1}$
   6a) Resample with weights $w_t^{(i)}$
   6b) Re-set weights $w_t^{(i)} = \frac{1}{N}$

III. Estimating the New Keynesian model

We show how to apply our estimation strategy to estimate the model of LS described in Section III.

A. The model and its state space representation

The model consists of equations (27), (28), (29) and (30).

In order to write the model in the Sims (2002) canonical form (A1) define $\eta_t^x = x_t - E_{t-1}(x_t)$, $\eta_t^\pi = \pi_t - E_{t-1}(\pi_t)$, $\xi_t^x = E_t(x_{t+1})$ and $\xi_t^\pi = E_t(\pi_{t+1})$. Then, the NK model can be expressed as:

$$\eta_t^x + \xi_{t-1}^x = \xi_t^x - \tau(R_t - \xi_t^\pi) + g_t \tag{A38}$$

$$\eta_t^\pi + \xi_{t-1}^\pi = \beta\xi_t^\pi + \kappa(\eta_t^x + \xi_{t-1}^x - z_t) \tag{A39}$$

$$R_t = \rho_R R_{t-1} + (1-\rho_R)(\psi_1\left(\eta_t^\pi + \xi_{t-1}^\pi\right) + \psi_2(\eta_t^x + \xi_{t-1}^x - z_t)) + \varepsilon_{R,t} \tag{A40}$$

Defining the vector $y_t = [\ x_t \quad \pi_t \quad R_t \quad \xi_t^x \quad \xi_t^\pi \quad g_t \quad z_t\ ]'$, the system in matrix form is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & (1-\rho_R)\psi_2 \\ 0 & 0 & -\tau & 1 & \tau & 1 & 0 \\ 0 & 0 & 0 & 0 & \beta & 0 & -\kappa \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \pi_t \\ R_t \\ \xi_t^x \\ \xi_t^\pi \\ g_t \\ z_t \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \rho_R & (1-\rho_R)\psi_2 & (1-\rho_R)\psi_1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\kappa & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_z \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \pi_{t-1} \\ R_{t-1} \\ \xi_{t-1}^x \\ \xi_{t-1}^\pi \\ g_{t-1} \\ z_{t-1} \end{bmatrix} +$$

$$+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{R,t} \\ \varepsilon_{g,t} \\ \varepsilon_{z,t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ (1-\rho_R)\psi_2 & (1-\rho_R)\psi_1 \\ 1 & 0 \\ -\kappa & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_t^x \\ \eta_t^\pi \end{bmatrix}$$

The class of solutions that we propose, parametrized by the matrix M_t, is written in equation (18) and it is expressed in terms of the vector $\tilde{l}_t = \begin{bmatrix} y_t \\ y_t^B \end{bmatrix}$, where y_t^B describes the evolution of the variables in the backward-looking solution. Note that in the vector $\tilde{l}_t$ the exogenous state variables g_t and z_t appear twice, since their dynamics are independent of M_t. For practical purposes, it is convenient to rewrite the solution in terms of a vector l_t where each exogenous shock is reported only once. First, define the following vectors:

$$y_{1,t} = [\ x_t \quad \pi_t \quad R_t \quad \xi_t^x \quad \xi_t^\pi\ ]'; \qquad\qquad y_{2,t} = [g_t \quad z_t\ ]'.$$

The solution can be partitioned as

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{1,t}^B \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \tilde{G}_{1,t} & \tilde{G}_{2,t} & \tilde{G}_{3,t} & \tilde{G}_{4,t} \\ \mathbf{0} & \tilde{G}_{5,t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{G}_{6,t} & \tilde{G}_{7,t} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{G}_{5,t} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{1,t-1}^B \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \tilde{H}_{1,t} \\ \tilde{H}_{2,t} \\ \tilde{H}_{3,t} \\ \tilde{H}_{2,t} \end{bmatrix} \varepsilon_t$$

where $\varepsilon_t = [\ \varepsilon_{R,t} \quad \varepsilon\prime_{g,t} \quad \varepsilon_{z,t}\ ]$. The endogenous variables in $y_{1,t}$ depend on the entire vector $\tilde{l}_{t-1}$, while the same variables in the backward-looking solution only depend on the backward-looking components of $\tilde{l}_{t-1}$. The exogenous variables, instead, are described by their own dynamics. It is straightforward, then, to rewrite the solution as:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{1,t}^B \end{bmatrix} = \begin{bmatrix} \tilde{G}_{1,t} & \left(\tilde{G}_{2,t} + \tilde{G}_{4,t}\right) & \tilde{G}_{3,t} \\ \mathbf{0} & \tilde{G}_{5,t} & \mathbf{0} \\ \mathbf{0} & \tilde{G}_{7,t} & \tilde{G}_{6,t} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{1,t-1}^B \end{bmatrix} + \begin{bmatrix} \tilde{H}_{1,t} \\ \tilde{H}_{2,t} \\ \tilde{H}_{3,t} \end{bmatrix} \varepsilon_t$$

that is, using a compact notation:

$$l_t = G_t l_{t-1} + H_t \varepsilon_t \tag{A41}$$

that is the state equation of system (A14), where the latent vector is:

$$l_t = [\ x_t \quad \pi_t \quad R_t \quad \xi_t^x \quad \xi_t^\pi \quad g_t \quad z_t \quad x_t^B \quad \pi_t^B \quad R_t^B \quad \xi_t^{xB} \quad \xi_t^{\pi B}\ ]'.$$

The observation equation is:

$$D_t = c + F l_t \tag{A42}$$

where D_t is a column vector with output gap, inflation and interest rate,

$$c = \begin{bmatrix} 0 \\ \pi^* \\ \pi^* + r^* \end{bmatrix}, \quad \text{and} \quad F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

B. The parameters updated through the Liu and West filter

The set of parameters has two components: $\theta = (\theta_1, \theta_2)$, where θ_1 contains all the parameters of the model except the variances:[5]

$$\theta_1 = [\ \rho_g \quad \rho_z \quad \rho_R \quad \kappa \quad \psi_1 \quad \psi_2 \quad \tau^{-1} \quad \pi^* \quad r^* \quad \gamma\ ]'.$$

Define the vector ϕ as a transformation of the vector θ_1 such that every element has support from $-\infty$ to $+\infty$. In particular, we use the logit function for the parameters that can take values in $[-1\ 1]$, and the logarithm for the parameters with positive support:

$$\phi^{(i)} = \begin{bmatrix} h(\rho_g^{(i)}) \\ h(\rho_z^{(i)}) \\ h(\rho_R^{(i)}) \\ \log(\kappa^{(i)}) \\ \log(\psi_1^{(i)}) \\ \log(\psi_2^{(i)}) \\ \log(\tau^{-1(i)}) \\ \log(\pi^{*(i)}) \\ \log(r^{*(i)}) \\ h(\gamma^{(i)}) \end{bmatrix}$$

where h is the logit function.

Finally, the parameter a in equation (A19) is set equal to 0.99.

C. The multiplicative sunspots

The latent process $M_{1,t}$ is updated using its law of motion. Under the stable model M_S we distinguish two cases: if condition (31) is not satisfied, $M_{1,t}^{(i)}$ can vary over time and we sample its values from the Normal distribution:

$$N\left(M_{t-1}^{(i)}, \sigma_\zeta^{2(i)}\right).$$

In contrast, if the Taylor principle is respected we set it equal to zero, that is the value corresponding to the unique stable solution.

Under the unstable model M_U, we first verify that the indicator function in (9) is equal to

[5] The parameter γ is estimated only under the unstable model and it is not included in the vector θ_1 under the stable model M_S.

one. Then, with probability $\gamma^{(i)}$, we draw $M_{1,t}^{(i)}$ from the Normal distribution:

$$N\left(\frac{M_{t-1}^{(i)}}{\gamma^{(i)}}, \sigma_{\zeta}^{2(i)}\right)$$

while we set it equal to zero with probability $\left(1-\gamma^{(i)}\right)$.

D. The parameters updated through Particle Learning

The vector θ_2 collects all the error variances and covariances:

$$\theta_2 = [\ \sigma_R^2 \quad \sigma_\zeta^2 \quad \sigma_g^2 \quad \sigma_z^2 \quad \rho_{gz}\ ]'. \tag{A43}$$

We follow the Particle Learning approach by Carvalho et al. (2010). The latent processes and the parameters in $\theta_1^{(i)}$ are used to update a set of sufficient statistics $s_t^{(i)}$ that contains $T_R^{(i)}$, $T_\zeta^{(i)}$, $T_{gz}^{(i)}$, $n_{\zeta,t}^{(i)}$ and, where:

$$T_R^{(i)} = \sum_{j=1}^{t}\left(\varepsilon_{R,j}^{(i)}\right)^2; \qquad T_\zeta^{(i)} = \sum_{j=1}^{n_{\zeta,t}^{(i)}}\left(\varepsilon_{\zeta,j}^{(i)}\right)^2; \qquad T_{gz}^{(i)} = \sum_{j=1}^{t}\left(\begin{bmatrix}\varepsilon_{g,j}^{(i)}\\ \varepsilon_{z,j}^{(i)}\end{bmatrix}\begin{bmatrix}\varepsilon_{g,j}^{(i)} & \varepsilon_{z,j}^{(i)}\end{bmatrix}\right)$$

and $n_{\zeta,t}^{(i)}$ is the number of times $M_t^{(i)}$ has been drawn from a Normal distribution rather than being set equal to zero. The sufficient statistics are then used to update the posterior distributions of the parameters in θ_2, which are known analytically (up to a normalizing constant), given our assumptions on the prior distributions. In particular, we assume that the priors for σ_R^2 and σ_ζ^2 have an Inverse Gamma distribution defined, respectively, by shape parameters a_R and a_ζ, and rate parameters b_R and b_ζ.[6] Their posterior distributions are also Inverse Gamma:

$$\left(\sigma_R^{2(i)}|D_t\right) \sim IG\left(a_R + \frac{t}{2}, b_R + \frac{T_R^{(i)}}{2}\right)$$
$$\left(\sigma_\zeta^{2(i)}|D_t\right) \sim IG\left(a_\zeta + \frac{n_{\zeta,t}^{(i)}}{2}, b_\zeta + \frac{T_\zeta^{(i)}}{2}\right).$$

Since the shocks to supply and demand are correlated, we assume that the prior for σ_g^2, σ_z^2 and the covariance ρ_{gz} is an Inverse Wishart with 8 degrees of freedom and scale matrix Σ_0. Given new data at time t, we can draw these parameters from their posterior distribution:

$$\left(\Sigma_{gz}|D_t\right) \sim IW\left(\Sigma_0 + T_{gz}, 8+t\right).$$

[6]These hyperparameters are such that the prior means and variances for σ_R^2 and σ_ζ^2 are the ones reported in Table 1.

E. The model under determinacy and stochastic volatility

In section V we compare the models M_S and M_U with a case in which we impose determinacy, but at the same time we allow the standard deviations of the structural shocks to vary over time. In this case we set $M_{1,t} = 1$ for every t, and we explore only the parameter space such that condition (31) is satisfied.

To estimate this model, we use the same algorithm described above with some modifications. First the parameter vector θ is partitioned as:

$$\theta_1 = [\ \rho_g \quad \rho_z \quad \rho_R \quad \kappa \quad \psi_1 \quad \psi_2 \quad \tau^{-1} \quad \pi^* \quad r^* \quad \gamma \quad \rho_{gz}\ ]'$$

and

$$\theta_2 = [\ \delta_R^2 \quad \delta_g^2 \quad \delta_z^2\ ]'. \tag{A44}$$

The latent processes are l_t, with dynamics described by equation (A41), and

$$\bar{\sigma}_t = [\ \log \sigma_{R,t} \quad \log \sigma_{g,t} \quad \log \sigma_{z,t}\ ]' \tag{A45}$$

with dynamics described by equation (13).

We take advantage of analytical integration, in analogy with the estimation of model M_S and M_U: conditional on $\bar{\sigma}_t$ the state space model for l_t is linear and Gaussian. Then, we modify the weights for the first resampling defined in equation (A18) (point 2a in the algorithm):

$$\tilde{w}_t^{(i)} \propto w_{t-1}^{(i)} p\left(D_t | l_{t-1}^{(i)}, g_{\bar{\sigma}}(\bar{\sigma}_{t-1}^{(i)}), m_{t-1}^{(i)}, \theta_2^{(i)}\right) \tag{A46}$$

where

$$g_{\bar{\sigma}}(\bar{\sigma}_{t-1}^{(i)}) = E\left(\bar{\sigma}_t^{(i)} | \bar{\sigma}_{t-1}^{(i)}\right) = \bar{\sigma}_{t-1}^{(i)} \tag{A47}$$

Moreover, in the *propagation* step, we keep $M_t = 1$ and we propagate $\bar{\sigma}_t^{(i)}$ from the distribution implied by its law of motion (13) (point 3b in the algorithm). The distribution of the latent process $l_t^{(i)}$ is again Normal, with mean and covariance matrix computed through the Kalman recursion (A25) to (A28), appropriately modified.

Finally, the set of sufficient statistics $s_t^{(i)}$ contains the following variables:

$$T_R^{(i)} = \sum_{j=1}^{t} \left(\nu_{R,j}^{(i)}\right)^2; \qquad T_g^{(i)} = \sum_{j=1}^{t} \left(\nu_{g,j}^{(i)}\right)^2; \qquad T_z^{(i)} = \sum_{j=1}^{t} \left(\nu_{z,j}^{(i)}\right)^2.$$

These allow us to draw δ_R, δ_g and δ_z from their posterior distributions.

F. Computational details

We work with 500,000 particles: this number is big enough to guarantee that the filter explores well the parameter space and the support of the latent processes at any time t. However, as is clear from Figure 5, when the inference on ψ_1 switches to the indeterminacy region, we observe a reduction in the variance of the posterior distribution. In order to make sure that this change in the distribution reflects the likelihood implied by new data, and not a technical problem related to the filter, we increase the number of particles to 2,000,000 from 1972:IV to 1979:II.

The particles are distributed to 44 cores which run in parallel. We use a computer with two processors Intel Xeon E5-2699 v4. The algorithm takes approximately 90 minutes to estimate the first subsample.

REFERENCES

Blanchard, Olivier J., "Backward and Forward Solutions for Economies with Rational Expectations," *American Economic Review*, 1979, *69*, 114–118.

Carvalho, Carlos M., Michael S. Johannes, Hedibert F. Lopes, and Nicholas G. Polson, "Particle Learning and Smoothing," *Statistical Science*, 02 2010, *25* (1), 88–106.

Liu, Jane and Mike West, "Combined parameters and state estimation in simulation-based filtering," in Springer, ed., *Sequential Monte Carlo Methods in Practice*, New York: Arnaud Doucet and Nando de Freitas and Neil Gordon, 2001, chapter 10, pp. 197–223.

Pitt, Michael K. and Neil Shephard, "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 1999, *94* (446), 590–599.

Sims, Christopher A., "Solving Linear Rational Expectations Models," *Computational Economics*, 2002, *20(1-2)*, 1–20.