Online Appendix:
A New Time-Varying Parameter Autoregressive Model for U.S. Inflation Expectations

Markku Lanne
University of Helsinki and CREATES

Jani Luoto
University of Helsinki

Parts A and B of this appendix provide details of the MCMC algorithms used for the estimation of the parameters of (1) and (18), respectively. The efficiency of these MCMC algorithms is discussed in part C.

Part A: MCMC Algorithm

In this part, we describe the MCMC algorithm used for the estimation of the parameters of (1). The sampler involves sequential drawings from the full conditional posteriors of $\boldsymbol{\mu}$**,** $\mathbf{h}$**,** $\boldsymbol{\phi}$**,** $\boldsymbol{\varphi}$**,** and $\boldsymbol{\epsilon}^{+}$. As discussed in Section 3, we employ the precision-based sampling method of Chan and Jeliazkov (2009), which exploits sparse matrix algorithms to efficiently obtain draws from the full conditionals. It is our experience that convergence occurs rapidly when the zeros of the polynomials $\phi_t(B)$ and $\varphi_t(B^{-1})$ do not lie too close to the unit circle. In this case, further tuning of the sampler is not needed, and according to the results of Koop and Potter (2011), the sampler provides an accurate approximation for the restricted posteriors of $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$.

It is convenient to express the prior defined in (9)–(13) in matrix notation comparable to (8). To that end, we define the following $T \times T$ matrix:

$$\mathbf{H_h} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho_h & 1 & 0 & \cdots & 0 \\ 0 & -\rho_h & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho_h & 1 \end{pmatrix},$$

and denote by $\mathbf{H}_{\boldsymbol{\mu}}$ the $T \times T$ first difference matrix obtained by setting $\rho_h = 1$ in $\mathbf{H_h}$. We further denote by $\mathbf{H}_{\boldsymbol{\phi}}$ and $\mathbf{H}_{\boldsymbol{\varphi}}$ the $Tr \times Tr$ and $Ts \times Ts$ first difference matrices, respectively. Then, the prior distributions of the latent state variables $\boldsymbol{\mu}$, $\mathbf{h}$, $\boldsymbol{\phi}$, and $\boldsymbol{\varphi}$ can be written as follows:

$$\begin{aligned} \boldsymbol{\mu} &\sim N\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\mu}}\right) f\left(\boldsymbol{\mu}\right), \ \mathbf{h} \sim N\left(\boldsymbol{\delta}_{\mathbf{h}}, \boldsymbol{\Omega}_{\mathbf{h}}\right), \\ \boldsymbol{\phi} &\sim N\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\phi}}\right) 1\left(\boldsymbol{\phi}\right), \text{ and } \boldsymbol{\varphi} \sim N\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\varphi}}\right) 1\left(\boldsymbol{\varphi}\right), \end{aligned} \tag{A.1}$$

where $\boldsymbol{\Omega}_i^{-1} = \mathbf{H}_i' \mathbf{S}_i^{-1} \mathbf{H}_i$ $(i \in \{\boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}\})$, $\boldsymbol{\delta}_{\mathbf{h}} = \mathbf{H}_{\mathbf{h}}^{-1} \widetilde{\boldsymbol{\delta}}_{\mathbf{h}}$, and $\widetilde{\boldsymbol{\delta}}_{\mathbf{h}} = (\mu_h, \mu_h (1 - \rho_h), \ldots, \mu_h (1 - \rho_h))'$. The indicator functions $1(\boldsymbol{\phi})$ and $1(\boldsymbol{\varphi})$ equal unity in the stationary region defined by (2) and zero otherwise, and $f(\boldsymbol{\mu})$ involves the distribution functions from the truncated normal distributions in (9) (see (A.3) below). The covariance matrices of the vectors of the innovations of the state variables are collected in $\mathbf{S}_{\boldsymbol{\mu}} = \text{diag}\left(\sigma_{0\mu}^2, \sigma_\mu^2, \ldots, \sigma_\mu^2\right)$, $\mathbf{S_h} = \text{diag}(\sigma_h^2/(1 - \rho_h^2), \ \sigma_h^2, \ldots, \sigma_h^2)$, $\mathbf{S}_{\boldsymbol{\phi}} = \text{diag}(Q_{0\phi}, Q_\phi, \ldots, Q_\phi)$, and $\mathbf{S}_{\boldsymbol{\varphi}} = \text{diag}(Q_{0\varphi}, Q_\varphi, \ldots, Q_\varphi)$, where $\sigma_{0\mu}^2$ and $\sigma_h^2/(1 - \rho_h^2)$ are the variances, and $Q_{0\phi}$, and $Q_{0\varphi}$ the covariance matrices of the respective initial states.

By a change of variable, the likelihood function (the joint density of $\mathbf{y}$) defined by (8) can be written as

$$\begin{aligned} p\left(\mathbf{y} \left| \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^+ \right.\right) \ &\propto \ \left|\det\left(\mathbf{P}\right)\right| \prod_{t=1}^{T} \exp\left(h_t/2\right) \\ & \quad \times \exp\left[-\frac{1}{2}\left(\mathbf{y} - \widetilde{\boldsymbol{\tau}} - \boldsymbol{\mu}\right)' \mathbf{P}' \mathbf{S}_{\boldsymbol{\epsilon}}^{-1} \mathbf{P} \left(\mathbf{y} - \widetilde{\boldsymbol{\tau}} - \boldsymbol{\mu}\right)\right], \end{aligned} \tag{A.2}$$

where $\widetilde{\boldsymbol{\tau}}= \mathbf{P}^{-1}\boldsymbol{\tau}$, and we assume that $|\mathbf{P}| \neq 0$. The result in (8), in turn follows from the fact discussed in Subsection 2.2 that model (1) can be expressed in matrix notation as

$$\mathbf{Py} = \mathbf{P}\boldsymbol{\mu} + \boldsymbol{\tau} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N\left(0, \mathbf{S}_{\boldsymbol{\epsilon}}\right)$$

with

$$\boldsymbol{\tau} = \begin{pmatrix} \rho^b_{1,1} y_0 + \cdots + \rho^b_{1,r} y_{-r+1} \\ \vdots \\ \rho^b_{r,r} y_0 \\ 0 \\ \vdots \\ 0 \\ \rho^f_{T-s+1,s} y^*_{T+1} \\ \vdots \\ \rho^f_{T,1} y^*_{T+1} + \cdots + \rho^f_{T,s} y^*_{T+s} \end{pmatrix}$$

and

$$\mathbf{P} = \begin{pmatrix} \rho_{1,0} & -\rho^f_{1,1} & \cdots & -\rho^f_{1,s} & 0 & & \cdots & & 0 \\ -\rho^b_{2,1} & \rho_{2,0} & -\rho^f_{2,1} & \cdots & -\rho^f_{2,s} & & & & \\ \vdots & \ddots & \ddots & & & \ddots & \ddots & & \\ -\rho^b_{r,r} & \cdots & -\rho^b_{r,1} & \rho_{r,0} & -\rho^f_{r,1} & \cdots & -\rho^f_{r,s} & & \\ 0 & \ddots & & \ddots & \ddots & \ddots & & \ddots & 0 \\ & & -\rho^b_{T-s,r} & \cdots & -\rho^b_{T-s,1} & \rho_{T-s,0} & -\rho^f_{T-s,1} & \cdots & -\rho^f_{T-s,s} \\ \vdots & & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ & & & & -\rho^b_{T-1,r} & \cdots & -\rho^b_{T-1,1} & \rho_{T-1,0} & -\rho^f_{T-1,1} \\ 0 & & \cdots & & 0 & -\rho^b_{T,r} & \cdots & -\rho^b_{T,1} & \rho_{T,0} \end{pmatrix},$$

where $\phi_t\left(B\right)\varphi_t\left(B^{-1}\right) = \rho_{t,0} - \sum_{i=1}^{r} \rho^b_{t,i} B^i - \sum_{i=1}^{s} \rho^f_{t,i} B^{-i}$, and $y^*_t = (y_t - \mu_t)$.

Sampling trend inflation $\boldsymbol{\mu}$. The conditional posterior density of $\boldsymbol{\mu}$ $(a < \mu_t < b)$ (for $t = 1, \ldots, T$) is obtained by combining the prior density defined by (A.1) with the likelihood function (A.2):

$$p\left(\boldsymbol{\mu}\left|\mathbf{y},\mathbf{h},\boldsymbol{\phi},\boldsymbol{\varphi},\boldsymbol{\epsilon}^{+}\right.\right)\propto\exp\left[-\frac{1}{2}\left(\boldsymbol{\mu}-\overline{\boldsymbol{\mu}}\right)'\overline{\mathbf{V}}_{\boldsymbol{\mu}}^{-1}\left(\boldsymbol{\mu}-\overline{\boldsymbol{\mu}}\right)\right]f\left(\boldsymbol{\mu}\right),$$

with

$$f\left(\boldsymbol{\mu}\right)\propto\prod_{t=2}^{T}\left[\Phi\left(\frac{b-\mu_{t-1}}{\sigma_{\mu}}\right)-\Phi\left(\frac{a-\mu_{t-1}}{\sigma_{\mu}}\right)\right]^{-1},\tag{A.3}$$

where $\Phi\left(\cdot\right)$ is the cumulative distribution function of the standard normal distribution, and

$$\overline{\mathbf{V}}_{\boldsymbol{\mu}}^{-1}=\mathbf{H}_{\mu}'\mathbf{S}_{\boldsymbol{\mu}}^{-1}\mathbf{H}_{\mu}+\mathbf{P}'\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{P},\text{ and }\overline{\boldsymbol{\mu}}=\overline{\mathbf{V}}_{\boldsymbol{\mu}}\mathbf{P}'\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{P}\left(\mathbf{y}-\mathbf{P}^{-1}\boldsymbol{\tau}\right).$$

Following Chan et al. (2013), we use an ARMH step with proposal distribution $N(\overline{\boldsymbol{\mu}},\overline{\mathbf{V}}_{\boldsymbol{\mu}})$ in sampling from the conditional posterior of $\boldsymbol{\mu}$. For the initial state, we assume a truncated normal prior distribution: $\mu_1\sim TN\left(a,b;\mu_0,\sigma_{0\mu}^2\right)$, where μ_0 and $\sigma_{0\mu}^2$ are known hyperparameters.

Sampling AR coefficients $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$. The conditional posteriors of $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ can be obtained by multiplying the likelihood function (A.2) by their prior densities (defined by (A.1)). The resulting conditional posterior distributions are nonstandard, and we therefore simulate from them using independence-chain Metropolis-Hastings steps as described below.

Let us first consider sampling from the conditional posterior of $\boldsymbol{\phi}$. To obtain a good proposal distribution for $\boldsymbol{\phi}$, we take $\boldsymbol{\varphi}$ as given and express (1) in terms of $u_t=\varphi_t\left(B^{-1}\right)y_t^*$ where $y_t^*=y_t-\mu_t$ is the demeaned inflation:

$$u_t=\mathbf{U}_t\boldsymbol{\phi}_t+\epsilon_t,$$

where $\mathbf{U}_t=\left(\varphi_t\left(B^{-1}\right)y_{t-1}^*,\ldots,\varphi_t\left(B^{-1}\right)y_{t-r}^*\right)$. Stacking the previous system for $t=1,\ldots,T$, results in

$$\mathbf{u}=\mathbf{U}\boldsymbol{\phi}+\boldsymbol{\epsilon},\qquad\boldsymbol{\epsilon}\sim N\left(0,\mathbf{S}_{\boldsymbol{\epsilon}}\right)\tag{A.4}$$

where $\mathbf{u} = (u_1, \ldots, u_T)'$, and

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & & \\ & \ddots & \\ & & \mathbf{U}_T \end{bmatrix}.$$

Combining (A.4) with (A.1), and using standard generalized linear regression results, we obtain the proposal density

$$q\left(\boldsymbol{\phi}\left|\mathbf{u}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^{+}\right.\right) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\phi}-\overline{\boldsymbol{\phi}}\right)' \overline{\mathbf{V}}_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{\phi}-\overline{\boldsymbol{\phi}}\right)\right] 1\left(\boldsymbol{\phi}\right),$$

where

$$\overline{\mathbf{V}}_{\boldsymbol{\phi}}^{-1} = \boldsymbol{\Omega}_{\boldsymbol{\phi}}^{-1} + \mathbf{U}'\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{U},\ \overline{\boldsymbol{\phi}} = \overline{\mathbf{V}}_{\boldsymbol{\phi}}\mathbf{U}'\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{u}.$$

It is straightforward to verify that the product $|\det(\mathbf{P})|\, q\left(\boldsymbol{\phi}\left|\mathbf{u}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^{+}\right.\right)$ provides a kernel of the conditional posterior $p\left(\boldsymbol{\phi}\left|\mathbf{y}, \mathbf{h}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^{+}\right.\right)$ (obtained by combining the likelihood function (A.2) and the prior density $p\left(\boldsymbol{\phi}\left|Q_{\phi}\right.\right) \propto N\left(\boldsymbol{\phi}\left|\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\phi}}\right.\right) 1\left(\boldsymbol{\phi}\right)$). Therefore, we use $N\left(\overline{\boldsymbol{\phi}}, \overline{\mathbf{V}}_{\boldsymbol{\phi}}\right) 1\left(\boldsymbol{\phi}\right)$ as the proposal distribution of $\boldsymbol{\phi}$. Following Chib and Greenberg (1994), we draw from the truncated normal by sampling from $N\left(\overline{\boldsymbol{\phi}}, \overline{\mathbf{V}}_{\boldsymbol{\phi}}\right)$ until we obtain a draw that lies in the stationary region. The prior distribution of the initial state $\boldsymbol{\phi}_1$ is assumed multivariate normal $N\left(\boldsymbol{\phi}_0, Q_{0\phi}\right)$ (see, e.g., Durbin and Koopman (2002)).

Turning to $\boldsymbol{\varphi}$, we take $\boldsymbol{\phi}$ as given and express (1) in terms of $\upsilon_t = \phi_t\left(B\right) y_t^*$:

$$\upsilon_t = \mathbf{V}_t \boldsymbol{\varphi}_t + \epsilon_t,$$

where $\mathbf{V}_t = \left(\phi_t\left(B\right) y_{t+1}^*, \ldots, \phi_t\left(B\right) y_{t+s}^*\right)$. Now, by following exactly the same steps as in the case of $\boldsymbol{\phi}$, we obtain the following Gaussian proposal density for $\boldsymbol{\varphi}$:

$$q\left(\boldsymbol{\varphi}\left|\boldsymbol{\upsilon}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\epsilon}^{+}\right.\right) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\varphi}-\overline{\boldsymbol{\varphi}}\right)' \overline{\mathbf{V}}_{\boldsymbol{\varphi}}^{-1}\left(\boldsymbol{\varphi}-\overline{\boldsymbol{\varphi}}\right)\right] 1\left(\boldsymbol{\varphi}\right),$$

where

$$\overline{\mathbf{V}}_{\boldsymbol{\varphi}}^{-1} = \boldsymbol{\Omega}_{\boldsymbol{\varphi}}^{-1} + \mathbf{V}'\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{V},\ \overline{\boldsymbol{\varphi}} = \overline{\mathbf{V}}_{\boldsymbol{\varphi}}\mathbf{V}'\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\upsilon},$$

$\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_T)'$, and

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & & \\ & \ddots & \\ & & \mathbf{V}_T \end{bmatrix}.$$

Analogously to $\boldsymbol{\phi}$, it can be shown that the kernel of the posterior density of $\boldsymbol{\varphi}$ can be expressed as $|\det(\mathbf{P})|\, q\left(\boldsymbol{\varphi}\left|\boldsymbol{\upsilon}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\epsilon}^{+}\right.\right)$, and we simulate candidate draws for $\boldsymbol{\varphi}$ from $N\left(\overline{\boldsymbol{\varphi}}, \overline{\mathbf{V}}_{\boldsymbol{\varphi}}\right) 1(\boldsymbol{\varphi})$. In the case of a purely noncausal TVP-AR$(0, s)$ process, $|\det(\mathbf{P})| = 1$ and $\mathbf{V}_t = (\upsilon_{t+1}, \ldots, \upsilon_{t+s})$, and the conditional posterior distribution of $\boldsymbol{\varphi}$ is truncated normal: $N\left(\overline{\boldsymbol{\varphi}}, \overline{\mathbf{V}}_{\boldsymbol{\varphi}}\right) 1(\boldsymbol{\varphi})$. The prior distribution of the initial state $\boldsymbol{\varphi}_1$ is assumed multivariate normal $N(\boldsymbol{\varphi}_0, Q_{0\varphi})$.

Sampling stochastic volatilities $\mathbf{h}$. In drawing the stochastic volatilities, we use the method proposed by Kim at al. (1998). By squaring and taking logs of the elements of $\mathbf{P}\left(\mathbf{y} - \mathbf{P}^{-1}\boldsymbol{\tau} - \boldsymbol{\mu}\right) \equiv \widetilde{\mathbf{y}} = \boldsymbol{\epsilon}$ (cf. (7)), with $\boldsymbol{\epsilon} = \mathbf{S}_{\boldsymbol{\epsilon}}^{1/2}\boldsymbol{\eta}$, and $\boldsymbol{\eta} \sim N(\mathbf{0}, I_T)$, we obtain

$$\widetilde{y}_t^{*} = h_t + \eta_t^{*}, \text{ for } t = 1, \ldots, T \tag{A.5}$$

where $\widetilde{y}_t^{*} = \log(\widetilde{y}_t^2 + 0.001)$, and the innovations $\eta_t^{*} = \log \eta_t^2$ are distributed as $\log \chi^2(1)$.[1] Following Kim et al. (1998), we approximate the $\log \chi^2(1)$ distribution by a mixture of seven normal distributions. In particular, $\{\eta_t^{*} | s_t = j\} \sim N\left(m_j - 1.2704, \nu_j^2\right)$, $\Pr(s_t = j) = q_j$, $j = 1, \ldots, 7$, where the parameters q_j, m_j, and ν_j are chosen to optimize the approximation, and are given in Kim et al. (1998). With this approximation, the joint density of $\widetilde{\mathbf{y}}^{*} = (\widetilde{y}_1^{*}, \ldots, \widetilde{y}_T^{*})$ in (A.5) is by a change of variable seen to be Gaussian. Hence, the conditional posterior distribution of $\mathbf{h}$ obtained by combining the density of $\widetilde{\mathbf{y}}^{*}$ and the prior density $p(\mathbf{h} | \mu_h, \rho_h, \sigma_h) = N(\mathbf{h} | \boldsymbol{\delta}_{\mathbf{h}}, \boldsymbol{\Omega}_{\mathbf{h}})$, defined by (A.1) is also Gaussian. The indicator variables $\mathbf{s} = (s_1, \ldots, s_T)$, needed to calculate the mixture of normals, are drawn independently from the discrete distribution

$$\Pr\left(s_t = j \left| \mathbf{y}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^{+}\right.\right) \propto q_j \nu_j^{-1} \exp\left[-\frac{1}{2\nu_j^2}\left(\widetilde{y}_t^{*} - h_t - m_j + 1.2704\right)^2\right].$$

[1] The offset constant (0.001) is used to ensure the robustness of the estimation procedure.

The prior distribution for initial state h_0 is assumed normal $N\left(\mu_h, \sigma_h^2/\left(1-\rho_h^2\right)\right)$ (cf. (10)).

Sampling future errors $\boldsymbol{\epsilon}^+$. This step is based on the approximation (discussed after equation (6)) $\phi_{T+h}(B)\, y^*_{T+h} \approx \sum_{j=0}^{M-h} \beta_{T+h,j}\epsilon_{T+h+j}$, which is used recursively to map from $\boldsymbol{\epsilon}^+$ into $y^*_{T+1}, \ldots, y^*_{T+s}$. In what follows, we use these two parameterizations interchangeably and assume a mixture of normals prior for $\boldsymbol{\epsilon}^+$: $\boldsymbol{\epsilon}^+ \sim N\left(\mathbf{0}, \mathbf{S}_{\boldsymbol{\epsilon}^+}\right)$, where $\mathbf{S}_{\boldsymbol{\epsilon}^+} = \text{diag}(\exp(h_{T+1}), \ldots, \exp(h_{T+M}))$ (see (13)). Then, the conditional posterior of $\boldsymbol{\epsilon}^+$ is obtained by multiplying its prior density by the likelihood function (A.2):

$$p\left(\boldsymbol{\epsilon}^+ \left|\mathbf{y}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}\right.\right) \propto \exp\left[-\frac{1}{2}\left(\widetilde{\mathbf{y}}'\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}^{-1}\widetilde{\mathbf{y}} + \boldsymbol{\epsilon}^{+\prime}\mathbf{S}_{\boldsymbol{\epsilon}^+}^{-1}\boldsymbol{\epsilon}^{+\prime}\right)\right],$$

where $\widetilde{\mathbf{y}}$ was defined in the discussion preceding equation (A.5). Since the posterior distribution of $\boldsymbol{\epsilon}^+$ is non-standard, we use an ARMH step to simulate from it.

An efficient proposal distribution providing an accurate approximation to the conditional posterior can be obtained using the (approximate) long forward moving average presentation (4) for $t = 1, \ldots, T$:

$$\boldsymbol{\upsilon} \approx \mathbf{X}_\beta \boldsymbol{\epsilon}^+ + \mathbf{B}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N\left(0, \mathbf{S}_{\boldsymbol{\epsilon}}\right)$$

where

$$\mathbf{X}_\beta = \begin{bmatrix} \beta_{1,T} & \beta_{1,T+1} & \cdots & \beta_{1,T+M-1} \\ \beta_{2,T-1} & \beta_{2,T} & \cdots & \beta_{2,T+M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{T,1} & \beta_{T,2} & \cdots & \beta_{T,M} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & \beta_{1,1} & \cdots & \beta_{1,T-1} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \beta_{T-1,1} \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The approximation error can be made negligible by setting M sufficiently large. A change of variable from $\boldsymbol{\epsilon}$ to $\boldsymbol{\upsilon}$ yields the following conditional density kernel of $\boldsymbol{\upsilon}$:

$$p\left(\boldsymbol{\upsilon}\left|\boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^+\right.\right) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\upsilon} - \mathbf{X}_\beta\boldsymbol{\epsilon}^+\right)' \mathbf{B}^{-1\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{B}^{-1}\left(\boldsymbol{\upsilon} - \mathbf{X}_\beta\boldsymbol{\epsilon}^+\right)\right],$$

where we have used the result that $|\mathbf{B}| = 1$. Multiplying the previous density by the prior density $p\left(\boldsymbol{\epsilon}^+\right) = N\left(\boldsymbol{\epsilon}^+ \left|\mathbf{0}, \mathbf{S}_{\boldsymbol{\epsilon}^+}\right.\right)$ and rearranging, results in a Gaussian proposal density for $\boldsymbol{\epsilon}^+$:

$$q\left(\boldsymbol{\epsilon}^+ \left|\boldsymbol{\upsilon}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}\right.\right) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\epsilon}^+ - \overline{\boldsymbol{\epsilon}}^+\right)' \overline{\mathbf{V}}_{\boldsymbol{\epsilon}^+}^{-1}\left(\boldsymbol{\epsilon}^+ - \overline{\boldsymbol{\epsilon}}^+\right)\right]$$

where

$$\overline{\mathbf{V}}_{\boldsymbol{\epsilon}^+}^{-1} = \mathbf{S}_{\boldsymbol{\epsilon}^+}^{-1} + \mathbf{X}'_\beta \mathbf{B}^{-1\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{B}^{-1}\mathbf{X}_\beta, \text{ and } \overline{\boldsymbol{\epsilon}}^+ = \overline{\mathbf{V}}_{\boldsymbol{\epsilon}^+}\mathbf{X}'_\beta \mathbf{B}^{-1\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{B}^{-1}\boldsymbol{\upsilon}.$$

Because the Gaussian distribution has exponentially decaying tails, and the posterior distribution of ϵ^+ tends to have fat tails, we modify the previous Gaussian proposal density to obtain a distribution with heavier tails. In particular, we consider a multivariate t distribution with ν degrees of freedom, location vector $\overline{\boldsymbol{\epsilon}}^+$, and scale matrix $\overline{\mathbf{V}}_{\boldsymbol{\epsilon}^+}$. We set the degree of freedom parameter ν to 5.

The coefficient vectors $\boldsymbol{\phi}_{T+1}, \ldots, \boldsymbol{\phi}_{T+s}$ and $\boldsymbol{\varphi}_{T+1}, \ldots, \boldsymbol{\varphi}_{T+s}$, needed to map $\boldsymbol{\epsilon}^+$ to $y^*_{T+1}, \ldots, y^*_{T+s}$, are simulated from the random walk models in (11) and (12), respectively, and the log volatilities $h_{T+1}, \ldots, h_{T+M}$ are obtained using (10).

Sampling Prior Hyperparameters $\mu_h, \rho_h, \sigma^2_\mu, \sigma^2_h$, Q_ϕ, , *and* Q_φ. Because of the hierarchical prior structure in which the prior hyperparameters affect the data only through the respective latent variables, their conditional posterior distributions are obtained directly from (9)–(12). In particular, the conditional distributions of the covariance matrices of the innovations of the AR coefficients are inverse-Wishart distributions (see Cogley and Sargent (2005)):

$$\left(Q_\phi \left|\boldsymbol{\phi}\right.\right) \sim iW\left(\underline{\nu}_\phi + T - 1, \underline{S}_\phi + \sum_{t=2}^{T}\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)'\right),$$

and

$$\left(Q_\varphi \left|\boldsymbol{\varphi}\right.\right) \sim iW\left(\underline{\nu}_v + T - 1, \underline{S}_\varphi + \sum_{t=2}^{T}\left(\boldsymbol{\varphi}_t - \boldsymbol{\varphi}_{t-1}\right)\left(\boldsymbol{\varphi}_t - \boldsymbol{\varphi}_{t-1}\right)'\right),$$

from which Q_ϕ and Q_φ can be drawn.

The conditional distribution of the variance of the innovations of the log volatility is also standard. This variance parameter can be sampled from the following inverse-gamma distribution:

$$\left(\sigma^2_h \left|\mathbf{h}, \mu_h, \rho_h\right.\right) \sim iG\left(\overline{\nu}_h, \overline{S}_h\right)$$

where $\overline{\nu}_h = \underline{\nu}_h + T/2$, and

$$\begin{aligned}\overline{S}_h &= \underline{S}_h + \left(1-\rho_h^2\right)\left(h_1-\mu_h\right)^2/2 \\ &+ \textstyle\sum_{t=2}^{T}\left(h_t-\mu_h-\rho_h\left(h_{t-1}-\mu_h\right)\right)^2/2.\end{aligned}$$

As to the innovation variance σ_μ^2, the density kernel of its nonstandard conditional posterior distribution is

$$p\left(\sigma_\mu^2\,|\boldsymbol{\mu}\right) \propto iG\left(\sigma_\mu^2\left|\underline{\nu}_\mu + \frac{T-1}{2}, \underline{S}_\mu + \frac{1}{2}\sum_{t=2}^{T}\left(\mu_t-\mu_{t-1}\right)^2\right.\right) f\left(\boldsymbol{\mu}\right),$$

where $f\left(\boldsymbol{\mu}\right)$ is given in (A.3). Following Chan et al. (2013), we simulate from the conditional posterior of σ_μ^2 via an independence-chain Metropolis-Hastings step using the first component of the product on the right hand side of the above expression as a proposal density.

Finally, we employ Chan's (in press) algorithm for drawing from the joint conditional posterior of μ_h and ρ_h ($|\rho_h| < 1$).

Student's t error distribution. Following Lanne and Saikkonen (2011), and Lanne, Luoma, and Luoto (2012), as an alternative to the assumption $\epsilon \sim N\left(0, \mathbf{S}_\epsilon\right)$, we also entertain the case where the error term ϵ_t follows Student's t distribution with location parameter zero, scale parameter $\exp(h_t)$, and λ degrees of freedom. In this case, the likelihood function can be readily obtained by employing the following parameterization (see Geweke (2005)):

$$\epsilon_t = \exp(h_t/2)\widetilde{h}_t^{-1/2}\eta_t, \tag{A.6}$$

where η_t is a standard normal random variable, and $\lambda\widetilde{h}_t$ is chi-square distributed with λ degrees of freedom ($\chi^2\left(\lambda\right)$). Under this parameterization, we only need to replace $\mathbf{S}_\epsilon$ with $\widetilde{\mathbf{S}}_\epsilon = \mathrm{diag}(\exp(h_1)/\widetilde{h}_1, \ldots, \exp(h_T)/\widetilde{h}_T)$ in (A.2) to obtain the likelihood function. Thus, simulation from the joint posterior distribution of the unobservables under Student's t errors only involves expanding the above algorithm by two additional Gibbs steps (for $\widetilde{\mathbf{h}} = \left(\widetilde{h}_1, \ldots, \widetilde{h}_T\right)'$ and λ).

In order to facilitate sampling $\widetilde{\mathbf{h}}$ and λ, we introduce three modifications. First, we interpret the random vector $\widetilde{\mathbf{h}}$ as a vector of parameters with hierarchical priors $\lambda\widetilde{h}_t \sim \chi^2(\lambda)$ and $\lambda \sim Exp(\underline{\lambda})$, where $\underline{\lambda}$ is a hyperparameter. Second, we extend the basic prior setup in (13), by assuming Student's t marginal prior distributions for the elements of $\boldsymbol{\epsilon}^+$: $\epsilon_{T+m} \sim t(0, \exp(h_{T+m}); \underline{\lambda})$ $(m = 1, \ldots, M)$. Third, we cluster $\boldsymbol{\epsilon}^+$ into two blocks of size M_1 and M_2 such that $M_1, M_2 > 0$ and $M_1 + M_2 = M$, and then perform sampling one block at time by an ARMH step. Under Student's t distributed errors, this should improve the performance of the sampler because typically the dependence of the errors weakens over time such that after, say, M_1 periods it is negligible. Therefore, the previous proposal of $\boldsymbol{\epsilon}^+$ needs to be modified to sample only from the conditional posterior of the first block efficiently, while the remaining M_2 error terms can be conveniently simulated using their marginal prior distributions as a proposal distribution.

Following Geweke (2005), we obtain the kernel of the conditional density of $\widetilde{\mathbf{h}}$, by multiplying the expanded likelihood function (A.2) by the prior density $p\left(\widetilde{\mathbf{h}}\,|\lambda\right)$. The kernel can be written as

$$p\left(\widetilde{h}_t\left|\mathbf{y}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^+, \lambda\right.\right) \propto \widetilde{h}_t^{(\lambda-1)/2} \exp\left[-\frac{1}{2}\left[\lambda + \exp\left(-h_t\right)\epsilon_t^2\right]\widetilde{h}_t\right],$$

where $\boldsymbol{\epsilon} = \widetilde{\mathbf{y}}$, $\widetilde{\mathbf{y}}$ is defined in the discussion preceding (A.5), and the precision parameters $\widetilde{h}_1, \ldots, \widetilde{h}_T$ are conditionally independent. Thus, by the properties of the chi-squared density, $\widetilde{h}_1, \ldots, \widetilde{h}_T$ can be sampled directly using

$$\left(\left[\lambda + \exp\left(-h_t\right)\epsilon_t^2\right]\widetilde{h}_t\left|\mathbf{y}, \boldsymbol{\mu}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\epsilon}^+, \lambda\right.\right) \sim \chi^2\left(\lambda + 1\right). \tag{A.7}$$

Because of the hierarchical prior structure in which λ affects the data only through

$\widetilde{\mathbf{h}}$, the conditional posterior density of λ becomes

$$p\left(\lambda \middle| \widetilde{\mathbf{h}}\right) \propto \left[2^{\lambda/2}\Gamma\left(\lambda/2\right)\right]^{-T} \lambda^{\lambda T/2} \left(\prod_{t=1}^{T} \widetilde{h}_t^{(\lambda-2)/2}\right)$$
$$\exp\left[-\left(\frac{1}{\underline{\lambda}} + \frac{1}{2}\sum_{t=1}^{T}\widetilde{h}_t\right)\lambda\right]$$

(see Geweke (2005) and Lanne, Luoma, and Luoto (2012)). As a candidate distribution for λ we use a univariate normal distribution, with mean equal to the mode of λ and precision parameter equal to the negative of the second derivative of the log conditional posterior density, evaluated at the mode.

Part B: Sampling parameters ϕ and φ for the NKPC

In this part, we describe the estimation of the parameters of (18). The proposed algorithm is a straightforward extension of that in part A of this appendix. We assume the truncated normal prior densities $TN\left(\underline{\phi}, \underline{h}_{\phi}^{-1}; 0 \leq \phi < 1\right)$ and $TN\left(\underline{\varphi}, \underline{h}_{\varphi}^{-1}; 0 \leq \varphi < 1\right)$ for ϕ and φ, respectively.

We start by deriving the conditional posterior distribution of ϕ. To that end, we first define the $(T\times 1)$ vector $\mathbf{v} = (v_1, \ldots, v_T)'$, where $v_t = (1-\phi B)\, y_t^*$ and $y_t^* = y_t - \mu_t$ is the demeaned inflation, and then use the definition $\varphi_t\left(B^{-1}\right) \equiv \varphi_t^*\left(B^{-1}\right)\left(1-\varphi B^{-1}\right)$ to rewrite (18) in matrix notations:

$$\mathbf{P_v v} = \boldsymbol{\tau}_\mathbf{v} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N\left(0, \mathbf{S}_{\boldsymbol{\epsilon}}\right).$$

The non-zero elements of $\mathbf{P_v}$ are as in (7), and obtained from the products of the polynomials $\phi_t^*(z)$ and $\varphi_t\left(z^{-1}\right)$ for $t = 1, \ldots, T$. The $T \times 1$ vector $\boldsymbol{\tau}_\mathbf{v}$ contains the initial values $v_{-r+1}, \ldots, v_0$ and the post sample observations $v_{T+1}, \ldots, v_{T+s}$.

Based on the definition $v_t = (1-\phi B)\, y_t^*$, we obtain

$$\mathbf{y}^* = \widetilde{\boldsymbol{\tau}}_\mathbf{v} + \phi\mathbf{x}^* + \mathbf{P}_\mathbf{v}^{-1}\boldsymbol{\epsilon},$$

where $\mathbf{y}^* = \left(y_1^*, \ldots, y_T^*\right)'$, $\mathbf{x}^* = \left(y_0^*, \ldots, y_{T-1}^*\right)'$, $\widetilde{\boldsymbol{\tau}}_\mathbf{v} = \mathbf{P}_\mathbf{v}^{-1}\boldsymbol{\tau}_\mathbf{v}$, and we assume that $|\mathbf{P_v}| \neq 0$ (in the restricted TVP-AR(1,3) model considered in Section 5.2 $|\mathbf{P_v}| = 1$).

Thus, according to standard results on linear regression, the conditional posterior distribution of ϕ is

$$\left(\phi\left|\mathbf{y},\boldsymbol{\mu},\mathbf{h},\boldsymbol{\phi},\boldsymbol{\varphi},\boldsymbol{\epsilon}^{+},\varphi\right.\right)\sim TN\left(\overline{\phi},\overline{h}_{\phi}^{-1};0\leq\phi<1\right),$$

where

$$\overline{\phi}=\overline{h}_{\phi}^{-1}\left[\underline{h}_{\phi}\underline{\phi}+\mathbf{x}^{*\prime}\mathbf{P}^{\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{P}\left(\mathbf{y}^{*}-\widetilde{\boldsymbol{\tau}}_{\mathbf{v}}\right)\right],\text{ and }\overline{h}_{\phi}=\left[\underline{h}_{\phi}+\mathbf{x}^{*\prime}\mathbf{P}_{\mathbf{v}}^{\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{P}_{\mathbf{v}}\mathbf{x}^{*}\right].$$

Thus, ϕ can be simulated from the truncated normal distribution.

We now turn to the conditional posterior distribution of φ. Stacking (18) for $t=1,\ldots,T$, we obtain

$$\mathbf{P}_{\boldsymbol{\omega}}\boldsymbol{\omega}=\boldsymbol{\tau}_{\boldsymbol{\omega}}+\boldsymbol{\epsilon},\qquad\boldsymbol{\epsilon}\sim N\left(0,\mathbf{S}_{\boldsymbol{\epsilon}}\right),$$

where $\boldsymbol{\omega}=\left(\omega_{1},\ldots,\omega_{T}\right)^{\prime}$, $\omega_{t}=\left(1-\varphi B^{-1}\right)v_{t}$, and the $T\times T$ matrix $\mathbf{P}_{\boldsymbol{\omega}}$ is as $\mathbf{P}$ in (7), and its non-zero elements are now obtained from the products of the polynomials $\phi_{t}^{*}\left(z\right)$ and $\varphi_{t}^{*}\left(z^{-1}\right)$ for $t=1,\ldots,T$. The $\left(T\times1\right)$ vector $\boldsymbol{\tau}_{\boldsymbol{\omega}}$ contains the initial values $\omega_{t-r+1},\ldots,\omega_{0}$ and the post sample observations $\omega_{T+1},\ldots,\omega_{T+s}$. Standard linear regression results and the assumption that $\left|\mathbf{P}_{\boldsymbol{\omega}}\right|\neq0$ imply that

$$p\left(\varphi\left|\mathbf{y},\boldsymbol{\mu},\mathbf{h},\boldsymbol{\phi},\boldsymbol{\varphi},\boldsymbol{\epsilon}^{+},\phi\right.\right)\propto f\left(\varphi\right)TN\left(\varphi\left|\overline{\varphi},\overline{h}_{\varphi}^{-1};0\leq\varphi<1\right.\right)$$

with

$$\overline{\varphi}=\overline{h}_{\varphi}^{-1}\left[\underline{h}_{\varphi}\underline{\varphi}+\mathbf{x}_{\mathbf{v}}^{*\prime}\mathbf{P}_{\boldsymbol{\omega}}^{\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{P}_{\boldsymbol{\omega}}\mathbf{v}\right],\text{ and }\overline{h}_{\varphi}=\left[\underline{h}_{\varphi}+\mathbf{x}_{\mathbf{v}}^{*\prime}\mathbf{P}_{\boldsymbol{\omega}}^{\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{P}_{\boldsymbol{\omega}}\mathbf{x}_{\mathbf{v}}^{*}\right],$$

where $\mathbf{x}_{\mathbf{v}}^{*}=\left(v_{2},\ldots,v_{T+1}\right)^{\prime}$. The function

$$f\left(\varphi\right)=\exp\left[\boldsymbol{\tau}_{\boldsymbol{\omega}}^{\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\mathbf{v}-0.5\boldsymbol{\tau}_{\boldsymbol{\omega}}^{\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\tau}_{\boldsymbol{\omega}}-\varphi\mathbf{x}_{\mathbf{v}}^{*\prime}\mathbf{P}_{\boldsymbol{\omega}}^{\prime}\mathbf{S}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\tau}_{\boldsymbol{\omega}}\right].$$

Note that in the TVP-AR(1,3) model considered in Section 5.2, $\left|\mathbf{P}_{\boldsymbol{\omega}}\right|=1$. We employ an independence-chain Metropolis-Hastings step to simulate from the conditional posterior of φ, using the second component on its right hand side as a proposal density.

To simulate the time-varying parameters $\phi_{t1}^{*},\ldots,\phi_{t,r-1}^{*}$ and $\varphi_{t1}^{*},\ldots,\varphi_{t,s-1}^{*}$, we use the algorithm described in part A of this appendix (replacing y_{t}^{*} with ω_{t} ($t=$

$1, \ldots, T)$). This algorithm can also be directly applied to the trend inflation $\boldsymbol{\mu}$ and the log variances $\mathbf{h}$, making use of the definitions $\phi_t(B) \equiv \phi_t^*(B)(1-\phi B)$ and $\varphi_t(B^{-1}) \equiv \varphi_t^*(B^{-1})(1-\varphi B^{-1})$. To simulate $\boldsymbol{\epsilon}^+$ we use the mapping between them and the post sample observations $v_{T+1}, \ldots, v_{T+s}$, and calculate the predictive distribution of $y_{T+1}^*, \ldots, y_{T+h}^*$ recursively using $v_t = (1-\phi B)\, y_t^*$.

Part C: Efficiency of MCMC Algorithm

Following Primiceri (2005) and Chan et al. (2013), we use the inefficiency factors (IF) to monitor the efficiency of the MCMC algorithm. The IF is defined as $1+2\sum_{k=1}^{K}\rho_k$, where ρ_k is the kth sample autocorrelation of the chain, and K is set in such a way that the autocorrelation tapers off (we follow Primiceri (2005) and use 4 percent tapered windows). Almost independent draws from the posterior suggest that an algorithm is efficient. In this case the IF are around 1. Because of the very high dimension of the TVP models, the IF are typically reported to be much higher (see, e.g., Primiceri (2005), Chan and Strachan (2012), and Chan et al. (2013)), the averages of IF ranging from 2 to more than 100. We follow Chan et al. (2013) and report the 50th, 25th, and 75th percentiles of the IF for the parameters of the unrestricted TVP-AR(1,3) model and the TVP-AR(1,3) model under the NKPC restrictions discussed in the text. In particular, the summary statistics are provided for six sets of parameters, consisting of the mean, the coefficients of the lag and leads, the stochastic volatilities, the future errors, and all the time-invariant parameters.

In the case of the TVP-AR(1,3) model, the median (and the 25th and 75th percentiles) of the IF is 3.6 (1.5, 4.2) for the mean, 4.9 (3.8, 8.0) for the coefficients of lagged inflation, 13.4 (6.4, 16.6) for the rest of the AR coefficients, 5.5 (3.6, 6.9) for the stochastic volatilities, 1.5 (0.99, 2.0) for the future errors, and 35.8 (17.8, 66.0) for the time-invariant parameters. The corresponding statistics for the parameters of the

TVP-AR(1,3) model incorporating the NKPC restrictions are 5.2 (3.3, 6.9), 6.6 (5.7, 8.4), 7.4 (5.2, 9.0), 1.3 (1.0, 1.6), and 41.8 (26.7, 61.4), respectively. These values are relatively small and comparable to those in Chan et al. (2013) and Chan (in press).

References

Chan, Joshua C. C., and Rodney W. Strachan (2012) "Estimation in non-linear non-Gaussian state-space models with precision-based methods," Technical report, Research School of Economics, Australian National University, Canberra.

Durbin, J. and S. J. Koopman (2002) "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis." *Biometrika*, 89, 603–616.

Geweke, John (2005) Contemporary Bayesian Econometrics and Statistics. Hoboken, NJ: Wiley.