Internet Appendix for
"Model Instability and Forecasting Performance"

Davide Pettenuzzo
Brandeis University*

Allan Timmermann
UCSD, CEPR, and CREATES†

May 5, 2015

A Posterior derivations

This not-for-publication appendix explains how we obtain parameter estimates for the models described in Section 2 of the paper, and shows how we use these to generate predictive densities.

A.1 Linear models

For the linear models the goal is to obtain draws from the joint posterior distribution $p\left(\mu, \boldsymbol{\beta}, \sigma_\varepsilon^{-2} \middle| \mathcal{Y}^t\right)$, where $\mathcal{Y}^t$ denotes all information available up to time t. Combining the priors in equations (2)-(4) of the paper with the likelihood function yields the following conditional posteriors:

$$\left[\begin{array}{c} \mu \\ \boldsymbol{\beta} \end{array}\right] \Bigg| \sigma_\varepsilon^{-2}, \mathcal{Y}^t \sim \mathcal{N}\left(\overline{\mathbf{b}}, \overline{\mathbf{V}}\right), \tag{A-1}$$

and

$$\sigma_\varepsilon^{-2} \Big| \mu, \boldsymbol{\beta}, \mathcal{Y}^t \sim \mathcal{G}\left(\overline{s}^{-2}, \overline{v}\right), \tag{A-2}$$

*Brandeis University, Sachar International Center, 415 South St, Waltham, MA, Tel: (781) 736-2834. Email: dpettenu@brandeis.edu.

†University of California, San Diego, 9500 Gilman Drive, MC 0553, La Jolla CA 92093. Tel: (858) 534-0894. Email: atimmerm@ucsd.edu.

where

$$\begin{aligned}
\overline{\mathbf{V}} &= \left[\underline{\mathbf{V}}^{-1} + \sigma_{\varepsilon}^{-2} \sum_{\tau=1}^{t-1} \mathbf{x}_{\tau} \mathbf{x}_{\tau}^{'} \right]^{-1}, \\
\overline{\mathbf{b}} &= \overline{\mathbf{V}} \left[\underline{\mathbf{V}}^{-1} \underline{\mathbf{b}} + \sigma_{\varepsilon}^{-2} \sum_{\tau=1}^{t-1} \mathbf{x}_{\tau} y_{\tau+1} \right], \\
\overline{v} &= \underline{v}_0 \left(t_0 - 1 \right) + \left(t - 1 \right).
\end{aligned} \tag{A-3}$$

and

$$\overline{s}^2 = \frac{\sum_{\tau=1}^{t-1} \left(y_{\tau+1} - \mu - \boldsymbol{\beta}' \mathbf{x}_{\tau} \right)^2 + \left(s_{y,t_0}^2 \times \underline{v}_0 \left(t_0 - 1 \right) \right)}{\overline{v}}. \tag{A-4}$$

A Gibbs sampler algorithm can be used to iterate back and forth between (A-1) and (A-2), yielding a series of draws for the parameter vector $(\mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2})$. Draws from the predictive density $p\left(y_{t+1} \middle| \mathcal{Y}^t \right)$ can then be obtained by noting that

$$p\left(y_{t+1} \middle| \mathcal{Y}^t \right) = \int p\left(y_{t+1} \middle| \mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2}, \mathcal{Y}^t \right) p\left(\mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2} \middle| \mathcal{Y}^t \right) d\mu d\boldsymbol{\beta} d\sigma_{\varepsilon}^{-2}. \tag{A-5}$$

Draws from $p\left(y_{t+1} \middle| \mathcal{Y}^t \right)$ are obtained in two steps:

1. Draw μ, $\boldsymbol{\beta}$, and $\sigma_{\varepsilon}^{-2}$ from $p\left(\mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2} \middle| \mathcal{Y}^t \right)$ using the Gibbs sampler described above

2. Given μ, $\boldsymbol{\beta}$, and $\sigma_{\varepsilon}^{-2}$, draw

$$y_{t+1} | \mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2}, \mathcal{Y}^t \sim \mathcal{N}\left(\mu + \boldsymbol{\beta}' \mathbf{x}_{\tau}, \sigma_{\varepsilon}^2 \right) \tag{A-6}$$

A.2 Time-varying Parameter, Stochastic Volatility Models

Let $\boldsymbol{\theta}_t$ be the time varying parameters, $\boldsymbol{\theta}_t = (\mu_t, \boldsymbol{\beta}_t')$, while $\boldsymbol{\theta}^t = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_t\}$ and $h^t = \{h_1, ..., h_t\}$ are the sequences of time-varying mean and log-volatility parameters up to time t. Finally, let $\Theta = \left(\mu, \boldsymbol{\beta}, \mathbf{Q}, \sigma_{\xi}^{-2}, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \lambda_0, \lambda_1 \right)$ be the time-invariant parameters of the TVP-SV model.

To obtain draws from the joint posterior distribution $p\left(\Theta, \boldsymbol{\theta}^t, h^t \middle| \mathcal{Y}^t \right)$ for the TVP-SV model, we use the Gibbs sampler to draw recursively from the following conditional distributions:[1]

1. $p\left(\boldsymbol{\theta}^t \middle| \Theta, h^t, \mathcal{Y}^t \right)$

2. $p\left(\mu, \boldsymbol{\beta} \middle| \Theta_{-\mu,\boldsymbol{\beta}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t \right)$

3. $p\left(\mathbf{Q} \middle| \Theta_{-\mathbf{Q}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t \right)$

[1] In standard set notation A_{-b} is the complementary set of b in A, i.e., $A_{-b} = \{x \in A : x \neq b\}$.

4. $p\left(h^t \middle| \Theta, \boldsymbol{\theta}^t, \mathcal{Y}^t\right)$

5. $p\left(\sigma_\xi^{-2} \middle| \Theta_{-\sigma_\xi^{-2}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t\right)$

6. $p\left(\boldsymbol{\gamma_\theta} \middle| \Theta_{-\boldsymbol{\gamma_\theta}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t\right)$

7. $p\left(\lambda_0, \lambda_1 \middle| \Theta_{-\lambda_0,\lambda_1}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t\right)$

We simulate from each of these blocks as follows. Starting with $\boldsymbol{\theta}^t$, we focus on $p\left(\boldsymbol{\theta}^t \middle| \Theta, h^t, \mathcal{Y}^t\right)$. First, define $\widetilde{y}_{\tau+1} = y_{\tau+1} - \mu - \boldsymbol{\beta}'\mathbf{x}_\tau$ and rewrite equation (6) in the paper as follows:

$$\widetilde{y}_{\tau+1} = \mu_{\tau+1} + \boldsymbol{\beta}'_{\tau+1} x_\tau + \exp\left(h_{\tau+1}\right) u_{\tau+1}. \tag{A-7}$$

Given a set of values for μ and $\boldsymbol{\beta}$, $\widetilde{y}_{\tau+1}$ is observable. This reduces (A-7) to the measurement equation of a standard linear Gaussian state space model with heteroskedastic errors. Thus the sequence of time varying parameters $\boldsymbol{\theta}^t$ can be drawn from (A-7) using the algorithm of Carter and Kohn (1994).

Second, conditional on $\boldsymbol{\theta}^t$ we can draw $\mu, \boldsymbol{\beta}$ from standard distributions for $p\left(\mu, \boldsymbol{\beta} \middle| \Theta_{-\mu,\boldsymbol{\beta}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t\right)$:

$$\left.\begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix}\right| \Theta_{-\mu,\boldsymbol{\beta}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t \sim N\left(\overline{\mathbf{b}}, \overline{\mathbf{V}}\right), \tag{A-8}$$

where

$$\begin{aligned} \overline{\mathbf{V}} &= \left[\underline{\mathbf{V}}^{-1} + \sum_{\tau=1}^{t-1} \frac{1}{\exp\left(h_{\tau+1}\right)^2} \mathbf{x}_\tau \mathbf{x}'_\tau\right]^{-1}, \\ \overline{\mathbf{b}} &= \overline{\mathbf{V}}\left[\underline{\mathbf{V}}^{-1}\underline{\mathbf{b}} + \sum_{\tau=1}^{t-1} \frac{1}{\exp\left(h_{\tau+1}\right)^2} \mathbf{x}_\tau \left(y_{\tau+1} - \mu_{\tau+1} - \boldsymbol{\beta}'_{\tau+1}\mathbf{x}_\tau\right)\right]. \end{aligned} \tag{A-9}$$

Third, note that

$$\mathbf{Q}| \Theta_{-\mathbf{Q}}, \theta^t, h^t, M'_i, \mathcal{Y}^t \sim \mathcal{IW}\left(\overline{\mathbf{Q}}, \overline{v}_{\mathbf{Q}}\right), \tag{A-10}$$

where

$$\overline{\mathbf{Q}} = \underline{\mathbf{Q}} + \sum_{\tau=1}^{t-1} \left(\boldsymbol{\theta}_{\tau+1} - \boldsymbol{\gamma}'_{\boldsymbol{\theta}}\boldsymbol{\theta}_\tau\right)\left(\boldsymbol{\theta}_{\tau+1} - \boldsymbol{\gamma}'_{\boldsymbol{\theta}}\boldsymbol{\theta}_\tau\right)'. \tag{A-11}$$

and $\overline{v}_{\mathbf{Q}} = (t-1) + \underline{v}_{\mathbf{Q}}\left(t_0 - 1\right)$. Fourth, define $y^*_{\tau+1} = y_{\tau+1} - (\mu + \mu_{\tau+1}) - \left(\boldsymbol{\beta} + \boldsymbol{\beta}_{\tau+1}\right)' \mathbf{x}_\tau$ and note that $y^*_{\tau+1}$ is observable conditional on μ, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}^t$. Next, rewrite equation (6) in the paper as

$$y^*_{\tau+1} = \exp\left(h_{\tau+1}\right) u_{\tau+1}. \tag{A-12}$$

Squaring and taking logs on both sides of (A-12) yields a new state space system that replaces equations (6)-(8) in the paper with

$$
\begin{aligned}
y^{**}_{\tau+1} &= 2h_{\tau+1} + u^{**}_{\tau+1}, \qquad \text{(A-13)}\\
h_{\tau+1} &= \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \qquad \text{(A-14)}
\end{aligned}
$$

where $y^{**}_{\tau+1} = \ln\left[\left(y^{*}_{\tau+1}\right)^2\right]$, and $u^{**}_{\tau+1} = \ln\left(u^2_{\tau+1}\right) \sim \ln\left(\chi^2_1\right)$, with u^{**}_{τ} independent of ξ_s for all τ and s. Kim et al. (1998) employ a data augmentation approach and introduce a new state variable s_τ, $\tau = 1, .., t$, turning their focus on drawing from $p\left(h^t \middle| \Theta, \boldsymbol{\theta}^t, s^t, \mathcal{Y}^t\right)$ instead of $p\left(h^t \middle| \Theta, \boldsymbol{\theta}^t, \mathcal{Y}^t\right)$.[2] Conditional on the additional state variable s_τ, the linear non-Gaussian state space representation in (A-13)-(A-14) can be written as an approximate linear Gaussian state space model:

$$
u^{**}_{\tau+1} \approx \sum_{j=1}^{7} q_j \mathcal{N}\left(m_j - 1.2704, v^2_j\right), \qquad \text{(A-15)}
$$

where m_j, v^2_j, and q_j, $j = 1, 2, ..., 7$, are constants specified in Kim et al. (1998). In turn, (A-15) implies

$$
u^{**}_{\tau+1} \middle| s_{\tau+1} = j \sim \mathcal{N}\left(m_j - 1.2704, v^2_j\right), \qquad \text{(A-16)}
$$

where $q_j = \Pr\left(s_{\tau+1} = j\right)$ is the probability of state j.

Conditional on s^t, we can rewrite the nonlinear state space system as follows:

$$
\begin{aligned}
y^{**}_{\tau+1} &= 2h_{\tau+1} + e_{\tau+1},\\
h_{\tau+1} &= \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \qquad \text{(A-17)}
\end{aligned}
$$

where $e_{\tau+1} \sim N\left(m_j - 1.2704, v^2_j\right)$ with probability q_j. We can use the algorithm of Carter and Kohn (1994) to draw the whole sequence of stochastic volatilities, h^t, for this linear Gaussian state space system.

Conditional on the sequence h^t, draws of states s^t can easily be obtained, noting that each of its elements can be independently drawn from the discrete density defined by

$$
\Pr\left(s_{\tau+1} = j \middle| y^{**}_{\tau+1}, h_{\tau+1}\right) = \frac{q_j f_{\mathcal{N}}\left(y^{**}_{\tau+1} \middle| 2h_{\tau+1} + m_j - 1.2704, v^2_j\right)}{\sum_{l=1}^{7} q_l f_{\mathcal{N}}\left(y^{**}_{\tau+1} \middle| 2h_{\tau+1} + m_l - 1.2704, v^2_l\right)}. \qquad \text{(A-18)}
$$

for $\tau = 1, ..., t-1$ and $j = 1, ..., 7$, and where $f_{\mathcal{N}}$ denotes the kernel of a normal density.

Fifth, the posterior distribution for $p\left(\sigma^{-2}_{\xi} \middle| \Theta_{-\sigma^{-2}_{\xi}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t\right)$ takes the form

$$
\sigma^{-2}_{\xi} \middle| \Theta_{-\sigma^{-2}_{\xi}}, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t \sim \mathcal{G}\left(\left[\frac{\sum_{\tau=1}^{t-1}\left(h_{\tau+1} - \lambda_0 - \lambda_1 h_\tau\right)^2 + \underline{k}_{\xi}\underline{v}_{\xi}\left(t-1\right)}{\left(t-1\right) + \underline{v}_{\xi}\left(t_0 - 1\right)}\right]^{-1}, \left(t-1\right) + \underline{v}_{\xi}\left(t_0 - 1\right)\right). \qquad \text{(A-19)}
$$

[2]Here $s^t = \{s_1, s_2, ..., s_t\}$ denotes the history up to time t of the new state variable s.

Sixth, obtaining draws from $p\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}\middle|\,\Theta_{-\boldsymbol{\gamma}_{\boldsymbol{\theta}}},\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\right)$ and $p\left(\lambda_0,\lambda_1\middle|\,\Theta_{-\lambda_0,\lambda_1},\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\right)$ is straightforward. As for $p\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}\middle|\,\Theta_{-\boldsymbol{\gamma}_{\boldsymbol{\theta}}},\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\right)$, we separately draw each of its elements. The $i-$th element $\gamma^{i}_{\boldsymbol{\theta}}$ is drawn from the following distribution

$$\gamma^{i}_{\boldsymbol{\theta}}\big|\,\Theta_{-\boldsymbol{\gamma}_{\boldsymbol{\theta}}},\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\sim\mathcal{N}\left(\overline{m}^{i}_{\gamma_{\boldsymbol{\theta}}},\overline{V}^{i}_{\gamma_{\boldsymbol{\theta}}}\right)\times\gamma^{i}_{\boldsymbol{\theta}}\in(-1,1) \tag{A-20}$$

where

$$\begin{aligned}\overline{V}^{i}_{\gamma_{\boldsymbol{\theta}}} &= \left[\underline{V}^{-1}_{\gamma_{\boldsymbol{\theta}}}+\mathbf{Q}^{ii}\sum_{\tau=1}^{t-1}\left(\theta^{i}_{\tau}\right)^{2}\right]^{-1}, \\ \overline{m}^{i}_{\gamma_{\boldsymbol{\theta}}} &= \overline{V}^{i}_{\gamma_{\boldsymbol{\theta}}}\left[\underline{V}^{-1}_{\gamma_{\boldsymbol{\theta}}}\underline{m}_{\gamma_{\boldsymbol{\theta}}}+\mathbf{Q}^{ii}\sum_{\tau=1}^{t-1}\theta^{i}_{\tau}\theta^{i}_{\tau+1}\right],\end{aligned} \tag{A-21}$$

and $\mathbf{Q}^{ii}$ is the $i-$th diagonal element of $\mathbf{Q}^{-1}$.

Finally, the distribution $p\left(\lambda_0,\lambda_1\middle|\,\Theta_{-\lambda_0,\lambda_1},\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\right)$ takes the form

$$\lambda_0,\lambda_1\big|\,\Theta_{-\lambda_0,\lambda_1},\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\sim\mathcal{N}\left(\begin{bmatrix}\overline{m}_{\lambda_0}\\ \overline{m}_{\lambda 1}\end{bmatrix},\overline{\mathbf{V}}_{\lambda}\right)\times\lambda_1\in(-1,1),$$

where

$$\overline{\mathbf{V}}_{\lambda}=\left\{\begin{bmatrix}\underline{V}^{-1}_{\lambda_0} & 0\\ 0 & \underline{V}^{-1}_{\lambda_1}\end{bmatrix}+\sigma_{\xi}^{-2}\sum_{\tau=1}^{t-1}\begin{bmatrix}1\\ h_{\tau}\end{bmatrix}[1,h_{\tau}]\right\}^{-1}, \tag{A-22}$$

and

$$\begin{bmatrix}\overline{m}_{\lambda_0}\\ \overline{m}_{\lambda 1}\end{bmatrix}=\overline{\mathbf{V}}_{\lambda}\left\{\begin{bmatrix}\underline{V}^{-1}_{\lambda_0} & 0\\ 0 & \underline{V}^{-1}_{\lambda_1}\end{bmatrix}\begin{bmatrix}\underline{m}_{\lambda_0}\\ \underline{m}_{\lambda_1}\end{bmatrix}+\sigma_{\xi}^{-2}\sum_{\tau=1}^{t-1}\begin{bmatrix}1\\ h_{\tau}\end{bmatrix}h_{\tau+1}\right\}. \tag{A-23}$$

Using these results, draws from the predictive density $p\left(y_{t+1}\middle|\,\mathcal{Y}^{t}\right)$ can be obtained by noting than

$$\begin{aligned}p\left(y_{t+1}\middle|\,\mathcal{Y}^{t}\right) &= \int p\left(y_{t+1}\middle|\,\boldsymbol{\theta}_{t+1},h_{t+1},\Theta,\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\right)\times p\left(\boldsymbol{\theta}_{t+1},h_{t+1}\middle|\,\Theta,\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\right) \\ &\quad\times p\left(\Theta,\boldsymbol{\theta}^{t},h^{t}\middle|\,\mathcal{Y}^{t}\right)d\Theta d\boldsymbol{\theta}^{t+1}dh^{t+1}.\end{aligned} \tag{A-24}$$

Draws from $p\left(y_{t+1}\middle|\,\mathcal{Y}^{t}\right)$ are obtained in three steps:

1. Draw from $p\left(\Theta,\boldsymbol{\theta}^{t},h^{t}\middle|\,\mathcal{Y}^{t}\right)$ using the above Gibbs sampling algorithm;

2. Simulate the future volatility, h_{t+1}, and the future regression coefficients, $\boldsymbol{\theta}_{t+1}$ from the distributions

$$h_{t+1}\big|\,\Theta,\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\sim\mathcal{N}\left(\lambda_0+\lambda_1 h_t,\sigma_{\xi}^{2}\right). \tag{A-25}$$

and

$$\boldsymbol{\theta}_{t+1}\big|\,\Theta,\boldsymbol{\theta}^{t},h^{t},\mathcal{Y}^{t}\sim\mathcal{N}\left(\boldsymbol{\gamma}'_{\boldsymbol{\theta}}\boldsymbol{\theta}_{t},\mathbf{Q}\right). \tag{A-26}$$

3. Finally, given $\boldsymbol{\theta}_{t+1}, h_{t+1}, \Theta, \mathcal{Y}^t$ draw

$$y_{t+1}|\, \boldsymbol{\theta}_{t+1}, h_{t+1}, \Theta, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t \sim \mathcal{N}\left((\mu + \mu_{t+1}) + \left(\boldsymbol{\beta} + \boldsymbol{\beta}_{t+1}\right)' \mathbf{x}_t, \exp\left(h_{t+1}\right)\right). \tag{A-27}$$

A.3 MS Models

To obtain draws from the joint posterior distribution $p\left(s^t, \Xi, \mathbf{P}|\, \mathcal{Y}^t\right)$ under the MS model, we use the Gibbs sampler to draw recursively from the following three conditional distributions:

1. $p\left(s^t|\, \Xi, \mathbf{P}, \mathcal{Y}^t\right)$
2. $p\left(\Xi|\, s^t, \mathbf{P}, \mathcal{Y}^t\right)$
3. $p\left(\mathbf{P}|\, s^t, \Xi, \mathcal{Y}^t\right)$

We simulate from each of these blocks as follows. We follow Chib (1996) and rely on a multi-move sampler for the path of hidden states, s^t. We first rewrite $p\left(s^t|\, \Xi, \mathbf{P}, \mathcal{Y}^t\right)$ as

$$p\left(s^t|\, \Xi, \mathbf{P}, \mathcal{Y}^t\right) = \left[\prod_{\tau=1}^{t-1} p\left(s_\tau|\, s_{\tau+1}, ..., s_t, \Xi, \mathbf{P}, \mathcal{Y}^t\right)\right] p\left(s_t|\, \Xi, \mathbf{P}, \mathcal{Y}^t\right). \tag{A-28}$$

$p\left(s_t|\, \Xi, \mathbf{P}, \mathcal{Y}^t\right)$ is the filtered probability distribution at $\tau = t$. Chib (1996) shows that

$$p\left(s_\tau|\, s_{\tau+1}, ..., s_t, \Xi, \mathbf{P}, \mathcal{Y}^t\right) \propto p\left(s_{\tau+1}|\, s_\tau, \mathbf{P}\right) \times p\left(s_\tau|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right), \tag{A-29}$$

where $p\left(s_\tau|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right)$ is the filtered probability distribution at τ, and $p\left(s_{\tau+1}|\, s_\tau, \mathbf{P}\right)$ is the transition probability from the Markov chain. Thus, to sample from $p\left(s^t|\, \Xi, \mathbf{P}, \mathcal{Y}^t\right)$, we first need to compute the sequence of filtered probability distributions $\left\{p\left(s_\tau|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right)\right\}_{\tau=1}^t$, which can be obtained by recursively iterating through the following two steps for $\tau = 1, 2, ..., t$:

$$p\left(s_\tau = l|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right) = \sum_{k=1}^{K} p\left(s_\tau = l|\, s_{\tau-1} = k, \mathbf{P}\right) p\left(s_{\tau-1} = k|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right), \tag{A-30}$$

and, for $l = 1, ..., K$,

$$p\left(s_\tau = l|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right) = \frac{p\left(y_\tau|\, s_\tau = l, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^{\tau-1}\right) p\left(s_\tau = l|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right)}{\sum_{k=1}^{K} p\left(y_\tau|\, s_\tau = k, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^{\tau-1}\right) p\left(s_\tau = k|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right)}. \tag{A-31}$$

At $\tau = 1$ the filter is started with the initial distribution $p\left(s_0|\, \mathbf{P}\right)$, which we set equal to the steady state probabilities. Once the sequence of filtered probabilities $\left\{p\left(s_\tau|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right)\right\}_{\tau=1}^t$

is available, we proceed as follows. First, we sample s_t from the filtered state probability distribution $p\left(s_t|\,\Xi,\mathbf{P},\mathcal{Y}^t\right)$. Next, for $\tau = t-1, t-2, ..., 1$ we sample s_τ from the conditional distribution $p\left(s_\tau = l|\,s_{\tau+1},...,s_t,\Xi,\mathbf{P},\mathcal{Y}^t\right)$

$$p\left(s_\tau = l|\,s_{\tau+1},...,s_t,\Xi,\mathbf{P},\mathcal{Y}^t\right) = \frac{p\left(s_{\tau+1} = l_m|\,s_\tau = l,\mathbf{P}\right)p\left(s_\tau = l|\,\Xi,\mathbf{P},\mathcal{Y}^\tau\right)}{\sum_{k=1}^K p\left(s_{\tau+1} = l_m|\,s_\tau = k,\mathbf{P}\right)p\left(s_\tau = k|\,\Xi,\mathbf{P},\mathcal{Y}^\tau\right)} \tag{A-32}$$

where l_m is the state drawn in the previous step of the recursion for $s_{\tau+1}$. Note that for each $\tau = t-1, t-2, ..., 1$, $p\left(s_\tau = l|\,s_{\tau+1},...,s_t,\Xi,\mathbf{P},\mathcal{Y}^t\right)$ needs to be evaluated for all $l = 1,..,K$.

The state-specific parameters $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K, \sigma_1^{-2}, ..., \sigma_K^{-2}$ are independent a posteriori and are drawn from the following distributions

$$\boldsymbol{\theta}_i|\,\sigma_i^{-2}, s^t, \mathbf{P}, \mathcal{Y}^t \sim \mathcal{N}\left(\overline{\mathbf{b}}_i, \overline{\mathbf{V}}_i\right), \tag{A-33}$$

and

$$\sigma_i^{-2}\Big|\,\boldsymbol{\theta}_i, s^t, \mathbf{P}, \mathcal{Y}^t \sim \mathcal{G}\left(\overline{s}_i^{-2}, \overline{v}_i\right), \tag{A-34}$$

where

$$\begin{aligned} \overline{\mathbf{V}}_i &= \left[\underline{\mathbf{V}}^{-1} + \sigma_i^{-2}\sum_{\tau:s_\tau=i}\mathbf{x}_\tau\mathbf{x}_\tau^{'}\right]^{-1}, \\ \overline{\mathbf{b}}_i &= \overline{\mathbf{V}}_i\left[\underline{\mathbf{V}}^{-1}\underline{\mathbf{b}} + \sigma_i^{-2}\sum_{\tau:s_\tau=i}\mathbf{x}_\tau y_{\tau+1}\right], \end{aligned} \tag{A-35}$$

and

$$\begin{aligned} \overline{v}_i &= \underline{v}_0 + n_i, \\ \overline{s}_i^2 &= \frac{\sum_{\tau:s_\tau=i}\left(y_{\tau+1} - \mu_i - \boldsymbol{\beta}_i'\mathbf{x}_\tau\right)^2 + \left(s_{y,t_0}^2 \times \underline{v}_0 n_i\right)}{\overline{v}_i}, \end{aligned} \tag{A-36}$$

where $n_i = \#\left(s_\tau = i\right)$ counts the number of observations from regime i along the path of hidden states s^t. To cope with the label switching problem that arises with Markov switching models, we identify different regimes by imposing the following constraint on the regime-specific volatilities: $\sigma_1^2 < \sigma_2^2 < ... < \sigma_K^2$.

Next, we draw the elements of the transition probability matrix $\mathbf{P}$ from $p\left(\mathbf{P}|\,s^t,\Xi,\mathcal{Y}^t\right)$. Because the rows $\mathbf{p}_{i,.}$ of $\mathbf{P}$ are independent a posteriori, we draw each row separately from the following Dirichlet distribution:

$$\mathbf{p}_{i,.}|\,s^t,\Xi,\mathcal{Y}^t \sim \mathcal{D}\left(e_{i1} + n_{i1}, ..., e_{iK} + n_{iK}\right), \quad i = 1, ..., K \tag{A-37}$$

where $n_{ij} = \#\left(s_{\tau-1} = i, s_\tau = j\right)$ counts the numbers of transitions from i to j as given by the whole path of hidden states s^t.

Finally, draws from the predictive density $p\left(y_{t+1}|\, \mathcal{Y}^t\right)$ can be obtained by noting than

$$\begin{aligned} p\left(y_{t+1}|\, \mathcal{Y}^t\right) &= \int p\left(y_{t+1}|\, s_{t+1}, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t\right) \times p\left(s_{t+1}|\, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t\right) \\ &\times p\left(s^t, \boldsymbol{\Xi}, \mathbf{P}\right|\, \mathcal{Y}^t\right) ds^{t+1} d\boldsymbol{\Xi} d\mathbf{P}. \end{aligned} \tag{A-38}$$

To draw from $p\left(y_{t+1}|\, \mathcal{Y}^t\right)$, we proceed in three steps:

1. Draw from $p\left(s^t, \boldsymbol{\Xi}, \mathbf{P}|\, \mathcal{Y}^t\right)$ using the above Gibbs sampling algorithm;

2. Simulate the time $t+1$ hidden state variable, s_{t+1} by drawing from $p\left(s_{t+1}|\, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t\right)$. Note that $p\left(s_{t+1}|\, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t\right)$ equals the $j-$th row of $\mathbf{P}$, $\mathbf{p}_{j,\cdot}$, if $s_t = j$;

3. Draw from $p\left(y_{t+1}|\, s_{t+1}, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t\right)$ using the distribution

$$y_{t+1}|\, s_{t+1}, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t \sim \mathcal{N}\left(\mu_{s_{t+1}} + \boldsymbol{\beta}'_{s_{t+1}}\mathbf{x}_t, \sigma^2_{s_{t+1}}\right). \tag{A-39}$$

A.4 CP Models

Draws from the joint posterior distribution $p\left(s^t, \Xi, \mathbf{P}|\, \mathcal{Y}^t\right)$ under the CP model are generated using a very similar set of steps as those used for the MS model. The key difference is of course the assumption of non-repeated regimes under the CP model. We follow Chib (1996) and Chib (1998) and rely on a multi-move sampler for the path of hidden states that is properly modified to deal with the constrained nature of the transition probability matrix $\mathbf{P}$. To sample from $p\left(s^t|\, \Xi, \mathbf{P}, \mathcal{Y}^t\right)$, we first compute the whole sequence of filtered probability distributions $\left\{p\left(s_\tau|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right)\right\}_{\tau=1}^t$, which can be obtained by iterating through the following two steps recursively for $\tau = 1, 2, ..., t$:

$$p\left(s_\tau = l|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right) = \sum_{k=l-1}^{l} p\left(s_\tau = l|\, s_{\tau-1} = k, \mathbf{P}\right) p\left(s_{\tau-1} = k|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right) \tag{A-40}$$

for $l = 1, ..., M$, and

$$p\left(s_\tau = l|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right) = \frac{p\left(y_\tau|\, s_\tau = l, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^{\tau-1}\right) p\left(s_\tau = l|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right)}{\sum_{k=l-1}^{l} p\left(y_\tau|\, s_\tau = k, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^{\tau-1}\right) p\left(s_\tau = k|\, \Xi, \mathbf{P}, \mathcal{Y}^{\tau-1}\right)} \tag{A-41}$$

Specifically, at $\tau = 1$ the filter is started by setting the initial distribution $p\left(s_0|\, \mathbf{P}\right)$ to a mass distribution that is concentrated at 1. Next, given the sequence of filtered probabilities $\left\{p\left(s_\tau|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right)\right\}_{\tau=1}^t$, we begin by setting $s_t = M$. Next, for $\tau = t-1, t-$

$2, ..., 1$ we sample s_τ from the conditional distribution $p\left(s_\tau = l|\, s_{\tau+1}, ..., s_t, \Xi, \mathbf{P}, \mathcal{Y}^t\right)$, given by

$$p\left(s_\tau = l|\, s_{\tau+1}, ..., s_t, \Xi, \mathbf{P}, \mathcal{Y}^t\right) = \frac{p\left(s_{\tau+1} = l_m|\, s_\tau = l, \mathbf{P}\right) p\left(s_\tau = l|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right)}{\sum_{k=l-1}^{l} p\left(s_{\tau+1} = l_m|\, s_\tau = k, \mathbf{P}\right) p\left(s_\tau = k|\, \Xi, \mathbf{P}, \mathcal{Y}^\tau\right)}, \tag{A-42}$$

where l_m is equal to the state drawn in the previous step of the recursion for $s_{\tau+1}$. The last of these distributions is degenerate at $s_1 = 1$. For a given sequence of states, the state-specific parameters $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_M, \sigma_1^{-2}, ..., \sigma_M^{-2}$ are drawn using similar steps as for the MS model.

To draw the elements of the transition probability matrix $\mathbf{P}$ from $p\left(\mathbf{P}|\, s^t, \Xi, \mathcal{Y}^t\right)$, note that the diagonal elements of $\mathbf{P}$ are independent a posteriori. We draw each of these separately from a Beta distribution

$$p_{ii}|\, s^t, \Xi, \mathcal{Y}^t \sim \mathcal{B}\left(\underline{a}_p + n_{ii}, \underline{b}_p + 1\right), \quad i = 1, ..., M-1, \tag{A-43}$$

where $n_{ii} = \#\left(s_{\tau-1} = i, s_\tau = i\right)$ counts the numbers of transitions from i to i observed on the path of hidden states s^t, and $n_{ii+1} = 1$ by construction.

Draws from the predictive density $p\left(y_{t+1}|\, s_{t+1} = M, \mathcal{Y}^t\right)$ are obtained by conditioning on no breaks between the end of the sample, t, and the end of the forecasting horizon, $t+1$. These are given by

$$p\left(y_{t+1}|\, s_{t+1} = M, \mathcal{Y}^t\right) = \int p\left(y_{t+1}|\, s_{t+1} = M, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t\right) \times p\left(s^t, \boldsymbol{\Xi}, \mathbf{P}\middle|\, \mathcal{Y}^t\right) ds^t d\boldsymbol{\Xi} d\mathbf{P}. \tag{A-44}$$

To obtain draws for $p\left(y_{t+1}|\, \mathcal{Y}^t\right)$, we proceed in two steps:

1. Draw from $p\left(s^t, \boldsymbol{\Xi}, \mathbf{P}|\, \mathcal{Y}^t\right)$ using the above Gibbs sampling algorithm;

2. Draw from $p\left(y_{t+1}|\, s_{t+1} = M, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t\right)$ using the distribution

$$y_{t+1}|\, s_{t+1} = M, s^t, \boldsymbol{\Xi}, \mathbf{P}, \mathcal{Y}^t \sim \mathcal{N}\left(\mu_M + \boldsymbol{\beta}_M' \mathbf{x}_t, \sigma_M^2\right). \tag{A-45}$$

References

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika 81*(3), pp. 541–553.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics 75*(1), 79 – 97.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*(2), 221 – 241.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies 65*(3), 361–393.