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 \( p(\mu, \beta, \sigma^{-2}_\varepsilon|\mathcal{Y}^t) \), 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:

\[
\begin{bmatrix} \mu \\ \beta \end{bmatrix} \mid \sigma^{-2}_\varepsilon, \mathcal{Y}^t \sim \mathcal{N}(\overline{b}, \overline{V}),
\]

(A-1)

and

\[
\sigma^{-2}_\varepsilon \mid \mu, \beta, \mathcal{Y}^t \sim \mathcal{G}(\overline{\sigma}^{-2}, \overline{v}),
\]

(A-2)
where

\[ \bar{V} = \left[ V^{-1} + \sigma^2 \sum_{\tau=1}^{t-1} x_\tau x_\tau' \right]^{-1}, \]

\[ \bar{b} = \bar{V} \left[ V^{-1} b + \sigma^2 \sum_{\tau=1}^{t-1} x_\tau y_{\tau+1} \right], \quad \bar{v} = v_0 (t_0 - 1) + (t - 1). \quad \text{(A-3)} \]

and

\[ s^2 = \sum_{\tau=1}^{t-1} (y_{\tau+1} - \mu - \beta' x_\tau)^2 + (s_{g,t_0}^2 \times v_0 (t_0 - 1)). \quad \text{(A-4)} \]

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 \((\mu, \beta, \sigma^{-2})\). Draws from the predictive density \(p(y_{t+1}|Y^t)\) can then be obtained by noting that

\[ p(y_{t+1}|Y^t) = \int p(y_{t+1}|\mu, \beta, \sigma^{-2}, Y^t) p(\mu, \beta, \sigma^{-2}|Y^t) d\mu d\beta d\sigma^{-2}. \quad \text{(A-5)} \]

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

1. Draw \(\mu, \beta, \) and \(\sigma^{-2}\) from \(p(\mu, \beta, \sigma^{-2}|Y^t)\) using the Gibbs sampler described above
2. Given \(\mu, \beta, \) and \(\sigma^{-2},\) draw

\[ y_{t+1}|\mu, \beta, \sigma^{-2}, Y^t \sim \mathcal{N}(\mu + \beta' x_\tau, \sigma^2). \quad \text{(A-6)} \]

**A.2 Time-varying Parameter, Stochastic Volatility Models**

Let \(\theta_t\) be the time varying parameters, \(\theta_t = (\mu_t, \beta'_t)\), while \(\theta^t = \{\theta_1, ..., \theta_t\}\) and \(h^t = \{h_1, ..., h_t\}\) are the sequences of time-varying mean and log-volatility parameters up to time \(t\). Finally, let \(\Theta = (\mu, \beta, Q, \sigma^{-2}, \gamma_\theta, \lambda_0, \lambda_1)\) be the time-invariant parameters of the TVP-SV model.

To obtain draws from the joint posterior distribution \(p(\Theta, \theta^t, h^t|Y^t)\) for the TVP-SV model, we use the Gibbs sampler to draw recursively from the following conditional distributions:\(^1\)

1. \(p(\theta^t|\Theta, h^t, Y^t)\)
2. \(p(\mu, \beta|\Theta_{-\mu, \beta}, \theta^t, h^t, Y^t)\)
3. \(p(Q|\Theta_{-Q}, \theta^t, h^t, Y^t)\)

---

\(^1\)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. \( p(h^t | \Theta, \theta^t, \mathcal{Y}^t) \)

5. \( p\left( \sigma^{-2}_\xi | \Theta_{-\sigma^{-2}_\xi}, \theta^t, h^t, \mathcal{Y}^t \right) \)

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

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

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

\[
\tilde{y}^t_{t+1} = \mu^t_{t+1} + \beta^t_{t+1} x^t_{t+1} + \exp\left( h^t_{t+1} \right) u^t_{t+1}.
\] (A-7)

Given a set of values for \( \mu \) and \( \beta \), \( \tilde{y}^t_{t+1} \) 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 \( \theta^t \) can be drawn from (A-7) using the algorithm of Carter and Kohn (1994).

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

\[
\begin{bmatrix} \mu \\ \beta \end{bmatrix} | \Theta_{-\mu, \beta}, \theta^t, h^t, \mathcal{Y}^t \sim N \left( \bar{b}, \bar{V} \right),
\] (A-8)

where

\[
\bar{V} = \left[ V^{-1} + \sum_{\tau=1}^{t-1} \frac{1}{\exp(h^t_{\tau+1})^2} x^t_{\tau} x'^t_{\tau} \right]^{-1},
\]

\[
\bar{b} = \bar{V} \left[ V^{-1} b + \sum_{\tau=1}^{t-1} \frac{1}{\exp(h^t_{\tau+1})^2} x^t_{\tau} \left( y^t_{\tau+1} - \mu^t_{\tau+1} - \beta^t_{\tau+1} x^t_{\tau} \right) \right].
\] (A-9)

Third, note that

\[
Q | \Theta_{-Q}, \theta^t, h^t, \mathcal{Y}^t \sim IW \left( \bar{Q}, \bar{v}_Q \right),
\] (A-10)

where

\[
\bar{Q} = Q + \sum_{\tau=1}^{t-1} \left( \theta^t_{\tau+1} - \gamma'_{\theta} \theta^t_{\tau} \right) \left( \theta^t_{\tau+1} - \gamma'_{\theta} \theta^t_{\tau} \right)'.
\] (A-11)

and \( \bar{v}_Q = (t - 1) + v_Q (t_0 - 1) \). Fourth, define \( y^*_t_{t+1} = y^t_{t+1} - (\mu + \mu^t_{t+1}) - (\beta + \beta^t_{t+1})' x^t_{t+1} \) and note that \( y^*_t_{t+1} \) is observable conditional on \( \mu, \beta, \) and \( \theta^t \). Next, rewrite equation (6) in the paper as

\[
y^*_t_{t+1} = \exp\left( h^t_{t+1} \right) u^t_{t+1}.
\] (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

\[
y_{\tau+1}^{**} = 2h_{\tau+1} + u_{\tau+1}^{**}, \\
h_{\tau+1} = \lambda_0 + \lambda_1 h_{\tau} + \xi_{\tau+1},
\]

(A-13)

(A-14)

where \( y_{\tau+1}^{**} = \ln \left( \left( y_{\tau+1}^{*} \right)^2 \right) \), and \( u_{\tau+1}^{**} = \ln \left( u_{\tau+1}^2 \right) \sim \ln(\chi_1^2) \), with \( u_{\tau}^{**} \) independent of \( \xi_s \) for all \( \tau \) and \( s \). Kim et al. (1998) employ a data augmentation approach and introduce a new state variable \( s_{\tau} \), \( \tau = 1, \ldots, t \), turning their focus on drawing from \( p\left(h^t|\Theta, \theta^t, s^t, \mathcal{Y}^t\right) \) instead of \( p\left(h^t|\Theta, \theta^t, s^t, \mathcal{Y}^t\right) \).^{2} Conditional on the additional state variable \( s_{\tau} \), the linear non-Gaussian state space representation in (A-13)-(A-14) can be written as an approximate linear Gaussian state space model:

\[
u_{\tau+1}^{**} \approx \sum_{j=1}^{7} q_j N\left(m_j - 1.2704, v_j^2\right),
\]

(A-15)

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

\[
u_{\tau+1}^{**} \mid s_{\tau+1} = j \sim N\left(m_j - 1.2704, v_j^2\right),
\]

(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:

\[
y_{\tau+1}^{**} = 2h_{\tau+1} + e_{\tau+1}, \\
h_{\tau+1} = \lambda_0 + \lambda_1 h_{\tau} + \xi_{\tau+1},
\]

(A-17)

where \( e_{\tau+1} \sim N\left(m_j - 1.2704, v_j^2\right) \) 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

\[
\Pr\left(s_{\tau+1} = j \mid y_{\tau+1}^{**}, h_{\tau+1}\right) = \frac{q_j f_N\left(y_{\tau+1}^{**} \mid 2h_{\tau+1} + m_j - 1.2704, v_j^2\right)}{\sum_{l=1}^{7} q_l f_N\left(y_{\tau+1}^{**} \mid 2h_{\tau+1} + m_l - 1.2704, v_l^2\right)}.
\]

(A-18)

for \( \tau = 1, \ldots, t - 1 \) and \( j = 1, \ldots, 7 \), and where \( f_N \) denotes the kernel of a normal density.

Fifth, the posterior distribution for \( p\left(\sigma^{-2}_\xi \mid \Theta^{-2}_{-\xi}, \theta^t, h^t, \mathcal{Y}^t\right) \) takes the form

\[
\sigma^{-2}_\xi \mid \Theta^{-2}_{-\xi}, \theta^t, h^t, \mathcal{Y}^t \sim \mathcal{G}\left(\left[\sum_{\tau=1}^{t-1} \frac{(h_{\tau+1} - \lambda_0 - \lambda_1 h_{\tau})^2 + k_\xi v_\xi (t - 1)}{(t - 1) + v_\xi (t_0 - 1)}\right]^{-1}, (t - 1) + v_\xi (t_0 - 1)\right).
\]

(A-19)

^{2}Here \( s^t = \{s_1, s_2, \ldots, s_t\} \) denotes the history up to time \( t \) of the new state variable \( s \).
Sixth, obtaining draws from $p(\gamma_\theta \mid \Theta - \gamma_\theta, \theta^t, h^t, y^t)$ and $p(\lambda_0, \lambda_1 \mid \Theta - \lambda_0, \lambda_1, \theta^t, h^t, y^t)$ is straightforward. As for $p(\gamma_\theta \mid \Theta - \gamma_\theta, \theta^t, h^t, y^t)$, we separately draw each of its elements. The $i$–th element $\gamma^i_\theta$ is drawn from the following distribution

$$
\gamma^i_\theta \mid \Theta - \gamma_\theta, \theta^t, h^t, y^t \sim N\left(\overline{m}^i_{\gamma_\theta}, \nabla^i_{\gamma_\theta}\right) \times \gamma^i_\theta \in (-1, 1) \quad (A-20)
$$

where

$$
\nabla^i_{\gamma_\theta} = \left[ V^{-1}_{\gamma_\theta} + Q^{ii} \sum_{\tau=1}^{t-1} (\theta^t)^2 \right]^{-1},
$$

$$
\overline{m}^i_{\gamma_\theta} = \nabla^i_{\gamma_\theta} \left[ V^{-1}_{\gamma_\theta} m_{\gamma_\theta} + Q^{ii} \sum_{\tau=1}^{t-1} \theta^t \theta^\tau_{t+1} \right], \quad (A-21)
$$

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

Finally, the distribution $p(\lambda_0, \lambda_1 \mid \Theta - \lambda_0, \lambda_1, \theta^t, h^t, y^t)$ takes the form

$$
\lambda_0, \lambda_1 \mid \Theta - \lambda_0, \lambda_1, \theta^t, h^t, y^t \sim N\left(\left[\begin{array}{c} \overline{m}_{\lambda_0} \\ \overline{m}_{\lambda_1} \end{array}\right], \nabla_\lambda\right) \times \lambda_1 \in (-1, 1),
$$

where

$$
\nabla_\lambda = \left\{ \left[ \begin{array}{cc} V^{-1}_{\lambda_0} & 0 \\ 0 & V^{-1}_{\lambda_1} \end{array} \right] + \sigma_{\xi}^{-2} \sum_{\tau=1}^{t-1} \left[ \begin{array}{c} 1 \\ h_{\tau} \end{array} \right] \right\}^{-1}, \quad (A-22)
$$

and

$$
\left[\begin{array}{c} \overline{m}_{\lambda_0} \\ \overline{m}_{\lambda_1} \end{array}\right] = \nabla_\lambda \left\{ \left[ \begin{array}{cc} V^{-1}_{\lambda_0} & 0 \\ 0 & V^{-1}_{\lambda_1} \end{array} \right] \left[\begin{array}{c} \overline{m}_{\lambda_0} \\ \overline{m}_{\lambda_1} \end{array}\right] + \sigma_{\xi}^{-2} \sum_{\tau=1}^{t-1} \left[ \begin{array}{c} 1 \\ h_{\tau} \end{array} \right] h_{\tau+1} \right\}. \quad (A-23)
$$

Using these results, draws from the predictive density $p(y_{t+1} \mid y^t)$ can be obtained by noting that

$$
p(y_{t+1} \mid y^t) = \int p(y_{t+1} \mid \theta_{t+1}, h_{t+1}, \Theta, \theta^t, h^t, y^t) \times p(\theta_{t+1}, h_{t+1} \mid \Theta, \theta^t, h^t, y^t) \, d\theta_{t+1} \, dh^t+1. \quad (A-24)
$$

Draws from $p(y_{t+1} \mid y^t)$ are obtained in three steps:

1. Draw from $p(\Theta, \theta^t, h^t \mid y^t)$ using the above Gibbs sampling algorithm;

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

$$
h_{t+1} \mid \Theta, \theta^t, h^t, y^t \sim N\left(\lambda_0 + \lambda_1 h_t, \sigma_{\xi}^2\right). \quad (A-25)
$$

and

$$
\theta_{t+1} \mid \Theta, \theta^t, h^t, y^t \sim N\left(\gamma_\theta \theta_t, Q\right). \quad (A-26)
$$
3. Finally, given \( \theta_{t+1}, h_{t+1}, \Theta, \mathcal{Y}^t \) draw

\[
y_{t+1} \mid \theta_{t+1}, h_{t+1}, \Theta, \theta', h', \mathcal{Y}^t \sim \mathcal{N}\left(\mu + \mu_{t+1} + (\beta + \beta_{t+1})'x_t, \exp(h_{t+1})\right).
\]

\[\text{(A-27)}\]

### A.3 MS Models

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

1. \( p(s^t \mid \Xi, P, \mathcal{Y}^t) \)
2. \( p(\Xi \mid s^t, P, \mathcal{Y}^t) \)
3. \( p(P \mid s^t, \Xi, \mathcal{Y}^t) \)

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 \( p(s^t \mid \Xi, P, \mathcal{Y}^t) \) as

\[
p\left(s^t \mid \Xi, P, \mathcal{Y}^t\right) = \left[ \prod_{t=1}^{t-1} p\left(s_{\tau} \mid s_{\tau+1}, \ldots, s_t, \Xi, P, \mathcal{Y}^t\right) \right] p\left(s_t \mid \Xi, P, \mathcal{Y}^t\right).
\]

\[\text{(A-28)}\]

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

\[
p\left(s_{\tau} \mid s_{\tau+1}, \ldots, s_t, \Xi, P, \mathcal{Y}^t\right) \propto p\left(s_{\tau+1} \mid s_{\tau}, P\right) \times p\left(s_{\tau} \mid \Xi, P, \mathcal{Y}^t\right),
\]

\[\text{(A-29)}\]

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

\[
p\left(s_{\tau} = l \mid \Xi, P, \mathcal{Y}_{\tau-1}\right) = \sum_{k=1}^K p\left(s_{\tau} = l \mid s_{\tau-1} = k, P\right) p\left(s_{\tau-1} = k \mid \Xi, P, \mathcal{Y}_{\tau-1}\right),
\]

\[\text{(A-30)}\]

and, for \( l = 1, \ldots, K \),

\[
p\left(s_{\tau} = l \mid \Xi, P, \mathcal{Y}_{\tau}\right) = \frac{p\left(y_{\tau} \mid s_{\tau} = l, \Xi, P, \mathcal{Y}_{\tau-1}\right) p\left(s_{\tau} = l \mid \Xi, P, \mathcal{Y}_{\tau-1}\right)}{\sum_{k=1}^K p\left(y_{\tau} \mid s_{\tau} = k, \Xi, P, \mathcal{Y}_{\tau-1}\right) p\left(s_{\tau} = k \mid \Xi, P, \mathcal{Y}_{\tau-1}\right)}.
\]

\[\text{(A-31)}\]

At \( \tau = 1 \) the filter is started with the initial distribution \( p(s_0 \mid P) \), which we set equal to the steady state probabilities. Once the sequence of filtered probabilities \( \{p(s_{\tau} \mid \Xi, P, \mathcal{Y}^t)\}_{\tau=1}^t \)
is available, we proceed as follows. First, we sample \( s_t \) from the filtered state probability distribution \( p(s_t | \Xi, P, Y^t) \). Next, for \( \tau = t-1, t-2, ..., 1 \) we sample \( s_\tau \) from the conditional distribution \( p(s_\tau = l | s_{\tau+1}, ..., s_t, \Xi, P, Y^t) \)

\[
p(s_\tau = l | s_{\tau+1}, ..., s_t, \Xi, P, Y^t) = \frac{p(s_{\tau+1} = l_m | s_\tau = l, P) p(s_\tau = l | \Xi, P, Y^\tau)}{\sum_{k=1}^{K} p(s_{\tau+1} = l_m | s_\tau = k, P) p(s_\tau = k | \Xi, P, Y^\tau)}
\]  

(A-32)

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

The state-specific parameters \( \theta_1, ..., \theta_K, \sigma_1^{-2}, ..., \sigma_K^{-2} \) are independent a posteriori and are drawn from the following distributions

\[
\theta_i | \sigma_i^{-2}, s^i, P, Y^t \sim N(\bar{b}_i, \bar{V}_i),
\]

(A-33)

and

\[
\sigma_i^{-2} | \theta_i, s^i, P, Y^t \sim G(\bar{s}_i^{-2}, \bar{v}_i),
\]

(A-34)

where

\[
\bar{V}_i = \left[ V^{-1} + \sigma_i^{-2} \sum_{\tau : s_\tau = i} x_\tau x'_\tau \right]^{-1},
\]

\[
\bar{b}_i = \bar{V}_i \left[ V^{-1} b + \sigma_i^{-2} \sum_{\tau : s_\tau = i} x_\tau y_{\tau+1} \right],
\]

(A-35)

and

\[
\bar{v}_i = v_{i0} + n_i,
\]

\[
\bar{s}_i^2 = \frac{\sum_{\tau : s_\tau = i} (y_{\tau+1} - \mu_i - \beta_i x_\tau)^2 + (s_{y,t0} \times v_{i0} n_i)}{\bar{v}_i},
\]

(A-36)

where \( n_i = \#(s_\tau = i) \) 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 \( P \) from \( p(P | s^t, \Xi, Y^t) \). Because the rows \( p_{i, \cdot} \) of \( P \) are independent a posteriori, we draw each row separately from the following Dirichlet distribution:

\[
p_{i, \cdot} | s^t, \Xi, Y^t \sim D(e_{i1} + n_{i1}, ..., e_{iK} + n_{iK}), \quad i = 1, ..., K
\]

(A-37)
where \( n_{ij} = \#(s_{r-1} = i, s_r = j) \) 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 \( p(y_{t+1}|\mathcal{Y}^t) \) can be obtained by noting that

\[
p(y_{t+1}|\mathcal{Y}^t) = \int p(y_{t+1}|s_{t+1}, s^t, \Xi, P, \mathcal{Y}^t) \times p(s_{t+1}|s^t, \Xi, P, \mathcal{Y}^t) \, ds_{t+1} d\Xi dP.
\]

(A-38)

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

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

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

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

\[
y_{t+1}|s_{t+1}, s^t, \Xi, P, \mathcal{Y}^t \sim \mathcal{N} \left( \mu_{s_{t+1}} + \beta'_{s_{t+1}} x_t, \sigma^2_{s_{t+1}} \right).
\]

(A-39)

### A.4 CP Models

Draws from the joint posterior distribution \( p(s^t, \Xi, P|\mathcal{Y}^t) \) 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 \( P \). To sample from \( p(s^t|\Xi, P, \mathcal{Y}^t) \), we first compute the whole sequence of filtered probability distributions \( \{p(s_r|\Xi, P, \mathcal{Y}^r)\}_{r=1}^t \), which can be obtained by iterating through the following two steps recursively for \( r = 1, 2, ..., t \):

\[
p(s_r = l|\Xi, P, \mathcal{Y}^{r-1}) = \sum_{k=l-1}^l p(s_r = l|s_{r-1} = k, P) p(s_{r-1} = k|\Xi, P, \mathcal{Y}^{r-1})
\]

(A-40)

for \( l = 1, ..., M \), and

\[
p(s_r = l|\Xi, P, \mathcal{Y}^r) = \frac{p(y_r|s_r = l, \Xi, P, \mathcal{Y}^{r-1}) p(s_r = l|\Xi, P, \mathcal{Y}^{r-1})}{\sum_{k=l-1}^l p(y_r|s_r = k, \Xi, P, \mathcal{Y}^{r-1}) p(s_r = k|\Xi, P, \mathcal{Y}^{r-1})}
\]

(A-41)

Specifically, at \( r = 1 \) the filter is started by setting the initial distribution \( p(s_0|P) \) to a mass distribution that is concentrated at 1. Next, given the sequence of filtered probabilities \( \{p(s_r|\Xi, P, \mathcal{Y}^r)\}_{r=1}^t \), we begin by setting \( s_t = M \). Next, for \( r = t-1, t-
2, ..., 1 we sample $s_\tau$ from the conditional distribution $p(s_\tau = l| s_{\tau + 1}, ..., s_t, \Xi, P, \mathcal{Y}^t)$, given by

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

(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 $\theta_1, ..., \theta_M, \sigma_1^{-2}, ..., \sigma_M^{-2}$ are drawn using similar steps as for the MS model.

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

$$p_{ii}| s^t, \Xi, \mathcal{Y}^t \sim B(a_p + n_{ii}, b_p + 1), \quad i = 1, ..., M - 1,$$

(A-43)

where $n_{ii} = \#(s_{\tau - 1} = i, s_\tau = i)$ 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 $p(y_{t+1}| s_{t+1} = M, \mathcal{Y}^t)$ 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

$$p(y_{t+1}| s_{t+1} = M, \mathcal{Y}^t) = \int p(y_{t+1}| s_{t+1} = M, s^t, \Xi, P, \mathcal{Y}^t) \times p(s^t, \Xi, P| \mathcal{Y}^t) \, ds^t d\Xi dP.$$

(A-44)

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

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

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

$$y_{t+1}| s_{t+1} = M, s^t, \Xi, P, \mathcal{Y}^t \sim N(\mu_M + \beta_M' x_t, \sigma_M^2).$$

(A-45)
References


