Hierarchical Markov Normal Mixture models with Applications to Financial Asset Returns Appendix: Proofs of Theorems and Conditional Posterior Distributions

John Geweke[a] and Gianni Amisano[b]

[a] *Departments of Economics and Statistics, University of Iowa, USA*

[b] *European Central Bank, Frankfurt, Germany and University of Brescia, Brescia, Italy*

December 22, 2008

Proofs of Theorems

Proof of Theorem 1

Using the methods of Ryden et al. (1998) for the Markov normal mixture model,

$$cov\left(y_t, y_{t-s} \mid \mathbf{x}_1, \ldots, \mathbf{x}_T\right) = \boldsymbol{\phi}' \, \mathbf{B}^{s'} \boldsymbol{\Pi} \, \boldsymbol{\phi} = \boldsymbol{\phi}' \boldsymbol{\Pi} \, \mathbf{B}^{s} \, \boldsymbol{\phi} \;\; (s = 1, 2, \ldots), \tag{1}$$

where $\mathbf{\Pi} = diag(\boldsymbol{\pi}), \mathbf{B} = \mathbf{P} - \mathbf{e}_{m_1}\boldsymbol{\pi}'$, which establishes sufficiency.

If the eigenvalues of $\mathbf{P}$ are distinct then $\mathbf{P}$ is diagonable and it has spectral decomposition $\mathbf{P} = \mathbf{Q}^{-1}\mathbf{\Lambda}\mathbf{Q}$, where the matrix $\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_{m_1}\right)$ contains the ordered eigenvalues λ_j of $\mathbf{P}$, $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \ldots |\lambda_{m_1}|$. The matrix $\mathbf{Q}$ has orthogonal columns and we may take

$$\begin{aligned} \mathbf{Q} &= \left[\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m\right]' = \left[\boldsymbol{\pi}, \mathbf{q}_2, \ldots, \mathbf{q}_m\right]', \\ \mathbf{Q}^{-1} &= \left[\mathbf{q}^1, \mathbf{q}^2, \ldots, \mathbf{q}^m\right] = \left[\mathbf{e}_m, \mathbf{q}^2, \ldots, \mathbf{q}^m\right]. \end{aligned}$$

If $\mathbf{P}$ is also irreducible and aperiodic then $\lambda_1 = 1 > |\lambda_2|$ and we may write

$$\mathbf{B} = \mathbf{Q}^{-1}\mathbf{\Lambda}\mathbf{Q} - \mathbf{q}^1\mathbf{q}_1' = \mathbf{Q}^{-1}\widetilde{\mathbf{\Lambda}}\mathbf{Q} \tag{2}$$

where $\widetilde{\mathbf{\Lambda}} = \mathrm{diag}\left(0, \lambda_2, \ldots, \lambda_{m_1}\right)$. From (1) absence of serial correlation is equivalent to

$$\boldsymbol{\phi}'\mathbf{\Pi}\mathbf{Q}^{-1}\widetilde{\mathbf{\Lambda}}^s\mathbf{Q}\boldsymbol{\phi} = 0 \quad (s = 1, 2, \ldots).$$

The first element of $\mathbf{Q}\boldsymbol{\phi}$ is $\mathbf{q}_1'\boldsymbol{\phi} = \boldsymbol{\pi}'\boldsymbol{\phi} = 0$, and so

$$\boldsymbol{\phi}'\mathbf{\Pi}\mathbf{Q}^{-1}\mathbf{\Lambda}^s\mathbf{Q}\boldsymbol{\phi} = 0 \quad (s = 1, 2, \ldots). \tag{3}$$

Define the $m_1 \times m_1$ matrix

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_{m_1} \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_{m_1}^2 \\ \vdots & \vdots & & \vdots \\ \lambda_1^{m_1} & \lambda_2^{m_1} & \cdots & \lambda_{m_1}^{m_1} \end{bmatrix},$$

whose determinant is $\left(\prod_{i=1}^{m_1} \lambda_i\right)^{m_1} \prod_{i<j} (\lambda_i - \lambda_j) \neq 0$ (Rao (1965), p 28). Let $\mathbf{A} = \mathbf{D}^{-1}$ and let $\delta_{i,j}$ denote the Kronecker delta function; then

$$\sum_{s=1}^{m_1} a_{is}\lambda_j^s = \delta_{i,j} \implies \sum_{i=1}^{m_1}\sum_{s=1}^{m_1} a_{is}\mathbf{\Lambda}^s = \mathbf{I}_{m_1} \;\; (i = 1, \ldots, m_1),$$

and from (3)

$$\sum_{i=1}^{m_1}\sum_{s=1}^{m_1} a_{is}\boldsymbol{\phi}'\boldsymbol{\Pi}\mathbf{Q}^{-1}\boldsymbol{\Lambda}^s\mathbf{Q}\boldsymbol{\phi} = \boldsymbol{\phi}'\boldsymbol{\Pi}\boldsymbol{\phi} = \sum_{i=1}^{m}\phi_i^2\pi_i = 0.$$

Proof of Theorem 2

The instantaneous variance matrix $\boldsymbol{\Gamma}_0^{(p)}$ is immediately attained by considering

$$\begin{aligned}
\boldsymbol{\Gamma}_0^{(p)} &= E\left[\mathbf{z}_t^{(p)} - \boldsymbol{\mu}^{*(p)}\right]\left[\mathbf{z}_t^{(p)} - \boldsymbol{\mu}^{*(p)}\right]' = E\left(\mathbf{z}_t^{(p)}\mathbf{z}_t^{(p)\prime}\right) - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{*(p)\prime} \\
&= \sum_{j=1}^{m}\pi_j\left[\mathbf{z}_t^{(p)}\mathbf{z}_t^{(p)\prime} \mid s_t = j\right] - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{*(p)\prime} \\
&= \sum_{j=1}^{m}\pi_j\left(\mathbf{R}_j^{(p)} + \boldsymbol{\mu}_j^{(p)}\boldsymbol{\mu}_m^{(p)\prime}\right) - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{*(p)\prime}.
\end{aligned}$$

The dynamic covariance matrices $\boldsymbol{\Gamma}_u^{(p)}$ $(p > 0)$ are obtained by conditioning on s_t and s_{t-u}, exploiting serial independence of observables after conditioning on the states, and then by marginalizing out the states:

$$\begin{aligned}
\boldsymbol{\Gamma}_u^{(p)} &= cov\left(\mathbf{z}_t^{(p)}, \mathbf{z}_{t-u}^{(p)}\right) = E\left(\mathbf{z}_t^{(p)}\mathbf{z}_{t-u}^{(p)\prime}\right) - \boldsymbol{\mu}^{*(p)}\boldsymbol{\mu}^{(p)\prime} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m} E\left(\mathbf{z}_t^{(p)}\mathbf{z}_{t-u}^{(p)\prime} \mid s_t = j, s_{t-u} = i\right)\left[\mathbf{P}^u\right]_{ij}\pi_i - \mathbf{M}^{(p)}\boldsymbol{\pi}\boldsymbol{\pi}'\mathbf{M}^{(p)\prime} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m} E\left(\mathbf{z}_t^{(p)} \mid s_t = j\right) E\left(\mathbf{z}_{t-u}^{(p)\prime} \mid s_t = i\right)\left[\mathbf{P}^u\right]_{ij}\pi_i - \mathbf{M}^{(p)}\boldsymbol{\pi}\boldsymbol{\pi}'\mathbf{M}^{(p)\prime} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m}\boldsymbol{\mu}_j^{(p)}\boldsymbol{\mu}_i^{(p)\prime}\left[\mathbf{P}^u\right]_{ij}\pi_i - \boldsymbol{\mu}^{(p)}\mathbf{e}_m'\boldsymbol{\Pi}\mathbf{M}^{(p)\prime} = \mathbf{M}^{(p)}\mathbf{B}^{u\prime}\boldsymbol{\Pi}\mathbf{M}^{(p)\prime},
\end{aligned}$$

where $\mathbf{B}^u = (\mathbf{P} - \mathbf{e}_m\boldsymbol{\pi}')^u = \mathbf{P}^u - \mathbf{e}_m\boldsymbol{\pi}'$.

Proof of Theorem 3

Adopt the notation in the proof of Theorem 2. From (2), $\mathbf{B}^u = \sum_{j=2}^m \lambda_j^u \mathbf{q}^j\mathbf{q}_j'$. Substituting in the expression for $\boldsymbol{\Gamma}_u^{(p)}$ in the statement of the theorem,

$$\boldsymbol{\Gamma}_u^{(p)} = \sum_{j=2}^{m}\lambda_j^u\mathbf{M}^{(p)}\mathbf{q}_j\mathbf{q}^{j\prime}\mathbf{M}^{(p)\prime} = \sum_{j=2}^{r+1}\lambda_j^u\mathbf{A}_j' \qquad (u = 1, 2, 3, \ldots)$$

where

$$\mathbf{A}_j' = \sum_{h\in H_j}\mathbf{M}^{(p)}\mathbf{q}_h\mathbf{q}^{h\prime}\mathbf{M}^{(p)\prime}, \;\; H_j = \left\{h : \mathbf{q}_h'\mathbf{P} = \lambda_j\mathbf{q}_h', \;\mathbf{M}^{(p)}\mathbf{q}_h \neq \mathbf{0}\right\}..$$

Observe that r is the number of distinct eigenvalues of $\mathbf{P}$ with modulus in the open unit interval associated with as least one column of $\mathbf{Q}'$ not in the column null space of $\mathbf{M}^{(p)}$. In other words, r can be less than $m-1$ because some eigenvalues are equal to zero (as in the compound Markov model interpreted as having $m = m_1 m_2$ states), because some eigenvalues are repeated, or because some eigenvalues are associated with columns of $\mathbf{Q}'$ all in the column null space of $\mathbf{M}^{(p)}$.

Define now a stochastic process $\mathbf{v}_t^{(p)}$ with autocovariances $\tilde{\mathbf{\Gamma}}_u^{(p)} = \sum_{j=2}^{r+1} \lambda_j^u \mathbf{A}_j'$ $(u > 0)$ and $\tilde{\mathbf{\Gamma}}_0^{(p)} = \sum_{j=2}^{r+1} \mathbf{A}_j'$. Then for $u > 0$, $\tilde{\mathbf{\Gamma}}_u^{(p)} = \mathbf{\Gamma}_u^{(p)}$, while

$$\tilde{\mathbf{\Gamma}}_0^{(p)} = \sum_{j=2}^{r+1} \mathbf{A}_j' = \sum_{j=1}^{m} \boldsymbol{\mu}_j^{(p)} \boldsymbol{\mu}_j'^{(p)} \pi_j - \boldsymbol{\mu}^{*(p)} \boldsymbol{\mu}^{*(p)\prime}.$$

Notice that the matrix $\mathbf{\Gamma}_0^{(p)} - \tilde{\mathbf{\Gamma}}_0^{(p)} = \sum_{j=1}^{m} \mathbf{R}_j^{(p)} \boldsymbol{\pi}_j$ is positive (semi) definite, since each $\mathbf{R}_j^{(p)}$ is a variance matrix.

Given that there are r distinct eigenvalues of $\mathbf{P}$, $\lambda_2, \ldots, \lambda_{r+1}$, with modulus in the open unit interval, contributing to the determination of $\mathbf{\Gamma}_u^{(p)} = \tilde{\mathbf{\Gamma}}_u^{(p)}$, there exists a unique set of constants $\alpha_1, \ldots, \alpha_r$ such that

$$\lambda_j^r - \sum_{i=1}^{r} \alpha_i \lambda_j^{r-i} = 0 \qquad (j = 2, \ldots, r+1)\,.$$

The coefficients $\alpha_1, \ldots, \alpha_r$ determine a degree r polynomial whose roots are $\lambda_2^{-1}, \ldots, \lambda_r^{-1}$. Thus for all $u > r$,

$$\begin{aligned}
\tilde{\mathbf{\Gamma}}_u^{(p)} - \sum_{i=1}^{r} \alpha_i \tilde{\mathbf{\Gamma}}_{u-i}^{(p)} &= \sum_{j=2}^{r+1} \lambda_j^u \mathbf{A}_j' - \sum_{i=1}^{r} \alpha_i \sum_{j=2}^{r+1} \lambda_j^{u-i} \mathbf{A}_j' \\
&= \sum_{j=2}^{r+1} \left(\lambda_j^u - \sum_{i=1}^{r} \alpha_i \lambda_j^{u-i} \right) \mathbf{A}_j' = \mathbf{0}.
\end{aligned}$$

The autocovariance function of $\left\{ \mathbf{v}_t^{(p)} \right\}$ therefore satisfies the Yule-Walker equations for a VAR(r) process with coefficient matrices $\alpha_i \mathbf{I}_{np} \quad (i = 1, \ldots, r)$.

Details of the Markov chain Monte Carlo algorithm

Let $\mathbf{s}^1 = (\mathbf{s}_{11}, \ldots, \mathbf{s}_{T1})'$. Then

$$\mathrm{p}\left(\mathbf{s}^1 \mid \mathbf{X}\right) = \pi_{\mathbf{s}_{11}} \prod_{t=2}^{T} p_{\mathbf{s}_{t-1,1}\mathbf{s}_{t1}} = \pi_{\mathbf{s}_{11}} \prod_{i=1}^{m_1} \prod_{j=1}^{m_1} p_{ij}^{T_{ij}}, \tag{4}$$

where T_{ij} is the number of transitions from persistent state i to j in $\mathbf{s}^1$. The $n \times n$ Markov transition matrix $\mathbf{P}$ is irreducible and aperiodic, and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{m_1})'$ is the unique stationary distribution of $\{s_{t1}\}$. Let $\mathbf{s}^2 = (s_{12}, \ldots, s_{T2})'$ denote all T transitory states. Then

$$p\left(\mathbf{s}^2 \mid \mathbf{s}^1, \mathbf{X}\right) = \prod_{t=1}^{T} \rho_{\mathbf{s}_t} = \prod_{i=1}^{m_1} \prod_{j=1}^{m_2} \rho_{ij}^{U_{ij}}. \tag{5}$$

where U_{ij} is the number of occurrences of $\mathbf{s}_t = (i, j)$ $(t = 1, \ldots, T)$.

The observables y_t depend on the latent states $\mathbf{s}_t$ and the deterministic variables $\mathbf{x}_t$. If $\mathbf{s}_t = (i, j)$ then

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \phi_i + \psi_{ij} + \varepsilon_t; \quad \varepsilon_t \sim N\left[\mathbf{0}, (h \cdot h_i \cdot h_{ij})^{-1}\right]. \tag{6}$$

Conditional on $(\mathbf{x}_t, \mathbf{s}_t)$ $(t = 1, \ldots, T)$ the y_t are independent. From (6) one expression for this distribution is

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{s}, \mathbf{X}) = {} & (2\pi)^{-Tn/2} h^{T/2} \prod_{i=1}^{m_1} h_i^{T_i n/2} \prod_{j=1}^{m_2} h_{ij}^{U_{ij} n/2} \\ & \cdot \exp\left[-h \sum_{i=1}^{m_1} h_i \sum_{j=1}^{m_2} h_{ij} \sum_{t:\mathbf{s}_t=(i,j)} \varepsilon_t^2/2\right], \end{aligned} \tag{7}$$

The unconditional mean of the transitory states within each permanent state is $\mathbf{0}$, which is equivalent to $\boldsymbol{\psi}_i'\boldsymbol{\rho}_i = \mathbf{0}$ $(i = 1, \ldots, m_1)$. Let $\mathbf{C}_j$ be an $m_2 \times (m_2 - 1)$ orthonormal complement of $\boldsymbol{\rho}_j$, define the $(m_2 - 1) \times 1$ vectors $\widetilde{\boldsymbol{\psi}}_j' = \mathbf{C}_j'\boldsymbol{\psi}_j$, and note that $\boldsymbol{\psi}_j = \mathbf{C}_j\widetilde{\boldsymbol{\psi}}_j$ $(j = 1, \ldots, m_1)$. Construct the $m_1 m_2 \times m_1 (m_2 - 1)$ block diagonal matrix $\mathbf{C} = \text{Blockdiag}\left[\mathbf{C}_1, \ldots, \mathbf{C}_{m_1}\right]$ and the $m_1 (m_2 - 1) \times 1$ vector $\widetilde{\boldsymbol{\psi}} = \left(\widetilde{\boldsymbol{\psi}}_1', \ldots, \widetilde{\boldsymbol{\psi}}_{m_1}'\right)'$. Then $\boldsymbol{\psi} = \mathbf{C}\widetilde{\boldsymbol{\psi}}$, and substituting in equation (7) at the end of Section 2.1.1,

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \widetilde{\boldsymbol{\phi}}'\mathbf{C}_0'\mathbf{z}_t^1 + \widetilde{\boldsymbol{\psi}}'\mathbf{C}'\mathbf{z}_t + \varepsilon_t. \tag{8}$$

This expression has the form $y_t = \boldsymbol{\gamma}'\mathbf{w}_t + \varepsilon_t$ in which the $(k + m_1 m_2 - 1) \times 1$ vector $\boldsymbol{\gamma} = \left(\boldsymbol{\beta}', \tilde{\boldsymbol{\phi}}', \widetilde{\boldsymbol{\psi}}'\right)'$ and

$$\mathbf{w}_t' = \left(\mathbf{x}_t', \mathbf{z}_t^{1\prime}\mathbf{C}_0, \mathbf{z}_t'\mathbf{C}\right). \tag{9}$$

Thus conditional on the latent states $\mathbf{s}_t$ (equivalently $\mathbf{z}_t^1$ and $\mathbf{z}_t^2$) $(t = 1, \ldots, T)$, and given the restrictions on the state means, (6) is a linear regression model with highly structured heteroscedasticity. If we take $\delta_t = h_{s_{t1}} h_{\mathbf{s}_t}$, then

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{s}, \mathbf{X}) &= (2\pi)^{-T/2} h^{T/2} \prod_{t=1}^{T} \delta_t^{n/2} \exp\left[-\sum_{t=1}^{T} h\delta_t \varepsilon_t^2/2\right] \\ &= (2\pi)^{-T/2} h^{T/2} \prod_{t=1}^{T} \delta_t^{n/2} \\ &\cdot \exp\left[-h \sum_{t=1}^{T} \delta_t \left(y_t - \mathbf{w}_t'\boldsymbol{\gamma}\right)^2 /2\right]. \end{aligned} \tag{10}$$

The kernel of the prior density is the product of the following expressions.

$$p(\boldsymbol{\beta}) \propto \exp\left[-\left(\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}\right)' \underline{\mathbf{H}}_{\boldsymbol{\beta}} \left(\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}\right) /2\right] \tag{11}$$

$$p(\mathbf{p}_i) \propto \prod_{j=1}^{m_1} p_{ij}^{r_1 - 1} \qquad (i = 1, \ldots, m_1) \tag{12}$$

$$p(\boldsymbol{\rho}_i) \propto \prod_{j=1}^{m_2} \rho_{ij}^{r_2 - 1} \qquad (i = 1, \ldots, m_1) \tag{13}$$

$$p(h) \propto h^{(\underline{\nu}-1)/2} \exp\left(-\underline{s}^2 h/2\right) \tag{14}$$

$$p(h_i) \propto h_i^{(\underline{\nu}_1-1)/2} \exp\left(-\underline{s}_1^2 h_i/2\right) \qquad (i = 1, \ldots, m_1) \tag{15}$$

$$\begin{aligned} p(h_{ij}) &\propto h_{ij}^{(\underline{\nu}_2-1)/2} \exp\left(-\underline{s}_2^2 h_{ij}/2\right) \\ &\quad (i = 1, \ldots, m_1;\ j = 1, \ldots, m_2) \end{aligned} \tag{16}$$

$$p\left(\widetilde{\boldsymbol{\phi}} \mid h\right) \propto h^{(m_1-1)/2} \exp\left(-\underline{h}_\phi h \widetilde{\boldsymbol{\phi}}'\widetilde{\boldsymbol{\phi}}/2\right) \tag{17}$$

$$= h^{(m_1-1)/2} \exp\left(-\underline{h}_\phi h \sum_{i=1}^{m_1-1} \widetilde{\phi}_i^2/2\right) \tag{18}$$

$$\begin{aligned} p\left(\widetilde{\boldsymbol{\psi}}_i \mid h_i, h\right) &\propto (h \cdot h_i)^{(m_2-1)/2} \\ &\cdot \exp\left(-\underline{h}_\psi h_i h \widetilde{\boldsymbol{\psi}}_i'\widetilde{\boldsymbol{\psi}}_i/2\right) \\ &\quad (i = 1, \ldots, m_1) \end{aligned}$$

$$
\begin{aligned}
& p\left(\widetilde{\boldsymbol{\psi}} \mid h_1, \ldots, h_m, h\right) \\
& \propto h^{m_1(m_2-1)/2} \prod_{i=1}^{m_1} h_i^{(m_2-1)/2} \exp\left(-\underline{h}_\psi h \sum_{i=1}^{m_1} h_i \widetilde{\boldsymbol{\psi}}_i' \widetilde{\boldsymbol{\psi}}_i / 2\right) \qquad (19) \\
& = h^{m_1(m_2-1)/2} \prod_{i=1}^{m_1} h_i^{(m_2-1)/2} \exp\left[-\underline{h}_\psi h \sum_{i=1}^{m_1} h_i \sum_{j=1}^{m_2-1} \widetilde{\psi}_{ij}^2 / 2\right] \qquad (20) \\
& = h^{m_1(m_2-1)/2} \prod_{i=1}^{m_1} h_i^{(m_2-1)/2} \\
& \cdot \exp\left\{-\underline{h}_\psi h \widetilde{\boldsymbol{\psi}}' \left[\operatorname{diag}\left(h_1, \ldots, h_{m_1}\right) \otimes \mathbf{I}_{m_2-1}\right] \widetilde{\boldsymbol{\psi}} / 2\right\} \qquad (21)
\end{aligned}
$$

Conditional posterior distribution of h. From (14), (18), (20) and (7),

$$
\begin{aligned}
& \overline{s}^2 h \sim \chi^2\left(\overline{\nu}\right); \\
& \overline{s}^2 = \underline{s}^2 + \zeta \underline{h}_\phi \widetilde{\boldsymbol{\phi}}' \widetilde{\boldsymbol{\phi}} + \underline{h}_\psi \sum_{i=1}^{m_1} h_i \widetilde{\boldsymbol{\psi}}_i' \widetilde{\boldsymbol{\psi}}_i + \sum_{t=1}^{T} \delta_t \varepsilon_t^2, \\
& \overline{\nu} = \underline{\nu} + \zeta\left(m_1 - 1\right) + m_1\left(m_2 - 1\right) + T.
\end{aligned}
$$

Conditional posterior distribution of the h_i. From (15), (20), and (7),

$$
\begin{aligned}
& \overline{s}_i^2 h_i \sim \chi^2\left(\overline{\nu}_i\right); \\
& \overline{s}_i^2 = \underline{s}_1^2 + \underline{h}_\psi h \widetilde{\boldsymbol{\psi}}_i' \widetilde{\boldsymbol{\psi}}_i + h \sum_{j=1}^{m_2} h_{ij} \sum_{t:\mathbf{s}_t=(i,j)} \varepsilon_t^2, \\
& \overline{\nu}_i = \underline{\nu}_1 + m_2 - 1 + nT_i
\end{aligned}
$$

$(i = 1, \ldots, m_1)$.

Conditional posterior distribution of the h_{ij}. From (16) and (7),

$$
\begin{aligned}
& \overline{s}_{ij}^2 h_{ij} \sim \chi^2\left(\overline{\nu}_{ij}\right); \\
& \overline{s}_{ij}^2 = \underline{s}_2^2 + h \cdot h_i \cdot \sum_{t:\mathbf{s}_t=(i,j)} \varepsilon_t^2, \\
& \overline{\nu}_{ij} = \underline{\nu}_2 + U_{ij}
\end{aligned}
$$

$(i = 1, \ldots, m_1;\ j = 1, \ldots, m_2)$.

Conditional posterior distribution of $\mathbf{P}$. From (12), (7), and (4),

$$
p\left(\mathbf{P}\right) \propto \pi_{s_{11}} \prod_{i=1}^{m_1} \prod_{j=1}^{m_1} p_{ij}^{r_1+T_{ij}-1} \exp\left(-h \sum_{t=1}^{T} \delta_t \varepsilon_t^2 / 2\right).
$$

Use a Metropolis within Gibbs step for each for each row i of $\mathbf{P}$. Draw the candidate $\mathbf{p}_i^* \sim \text{Beta}\left(r_1 + T_{i1}, \ldots, r_1 + T_{im_1}\right)$, and let $\mathbf{C}_0^*$ be the orthonormal complement of $\boldsymbol{\pi}^*$ corresponding to the resulting $\mathbf{P}^*$. Account must be taken of the fact that because $\varepsilon_t = y_t - \boldsymbol{\beta}'\mathbf{x}_t - \boldsymbol{\psi}_{\mathbf{s}_t} - \mathbf{z}_t^{1\prime}\mathbf{C}_0\widetilde{\boldsymbol{\phi}}$, $\mathbf{C}_0$ is a function of $\boldsymbol{\pi}$ and therefore of $\mathbf{P}$. Let $\mathbf{C}_0^*$ be the orthonormal complement of $\boldsymbol{\pi}^*$ and compute $\varepsilon_t^* = y_t - \boldsymbol{\beta}'\mathbf{x}_t - \boldsymbol{\psi}_{\mathbf{s}_t} - \mathbf{z}_t^{1\prime}\mathbf{C}_0^*\widetilde{\boldsymbol{\phi}}$. The Metropolis acceptance ratio is

$$\frac{\pi_{s_{11}}^* \exp\left(-\zeta h \sum_{t=1}^T \delta_t \varepsilon_t^{*2}/2\right)}{\pi_{s_{11}} \exp\left(-\zeta h \sum_{t=1}^T \delta_t \varepsilon_t^2/2\right)}.$$

If the candidate is accepted, then $\mathbf{P}$ is updated to $\mathbf{P}^*$, $\boldsymbol{\pi}$ to $\boldsymbol{\pi}^*$, and $\mathbf{C}_0$ to $\mathbf{C}_0^*$.

The orthonormal complement of $\mathbf{C}_0$ of $\boldsymbol{\pi}$ is not unique. As discussed in Section 2.1.2 nothing substantive in the model depends on which $\mathbf{C}_0$ is used. However, if $\mathbf{C}_0$ is not a smooth function of $\boldsymbol{\pi}$ then the candidate will be rejected more often than if it is, because $\mathbf{C}_0\widetilde{\boldsymbol{\phi}}$ will change more. To construct a unique orthonormal complement $\mathbf{C}$ that is a smooth function of a vector of probabilities $\boldsymbol{\pi}$ with $\sum_{i=1}^m \pi_i = 1$, note that $\pi_j \in (0,1)$ with probability 1 $(j = 1, \ldots, m)$. Construct a matrix $\mathbf{C}^*$ as follows. The first column of $\mathbf{C}^*$ is $c_{11}^* = \pi_2$, $c_{21}^* = -\pi_1$, $c_{i1}^* = 0 \;\; (i = 3, \ldots, m)$. The j'th column of $\mathbf{C}^*$ is $c_{ij}^* = \pi_i \;\; (i = 1, \ldots, j)$, $c_{j+1,j}^* = -\sum_{i=1}^j \pi_i^2/\pi_{j+1}$, $c_{ij}^* = 0 \;\; (i = j+2, \ldots, m)$. Construct $\mathbf{C}$ from $\mathbf{C}^*$ by normalizing the columns to each have Euclidian length 1.

Conditional posterior distribution of $\mathbf{R}$. From (13), (7), and (5),

$$p\left(\boldsymbol{\rho}_j\right) \propto \prod_{k=1}^{m_2} \rho_{jk}^{r_2 + U_{jk} - 1} \exp\left(-h \sum_{t: s_{t1} = j} \delta_t \varepsilon_t^2/2\right).$$

Use a Metropolis within Gibbs step for each for each row j of $\mathbf{R}$. Note that in $\varepsilon_t = y_t - \boldsymbol{\beta}'\mathbf{x}_t - \boldsymbol{\phi}_{\mathbf{s}_t} - \mathbf{z}_t'\mathbf{C}\widetilde{\boldsymbol{\psi}}$, $\mathbf{C}_j$ is a function of $\boldsymbol{\rho}_j$ whenever $s_{t1} = j$. Draw the candidate $\boldsymbol{\rho}_j^*$ from $\text{Beta}\left(r_2 + U_{j1}, \ldots, r_2 + U_{j,m_2}\right)$. Let $\mathbf{C}_j^*$ be the orthonormal complement of $\boldsymbol{\rho}_j^*$. For all t for which $s_{t1} = j$, compute $\varepsilon_t^* = y_t - \boldsymbol{\beta}'\mathbf{x}_t - \boldsymbol{\phi}_{\mathbf{s}_t} - \mathbf{z}_t'\mathbf{C}^*\widetilde{\boldsymbol{\psi}}$.The Metropolis acceptance ratio is

$$\frac{\exp\left(-h \sum_{t: s_{t1} = j} \delta_t \varepsilon_t^{*2}/2\right)}{\exp\left(-h \sum_{t: s_{t1} = j} \delta_t \varepsilon_t^2/2\right)}.$$

The Metropolis step is used only after the first 1,000 iterations.

Conditional posterior distribution of $\boldsymbol{\gamma}$. Recall that $y_t = \mathbf{w}_t'\boldsymbol{\gamma} + \varepsilon_t$, with $\boldsymbol{\gamma}' = \left(\boldsymbol{\beta}', \tilde{\boldsymbol{\phi}}', \tilde{\boldsymbol{\psi}}'\right)$ and

$$\mathbf{w}_t' = \left(\mathbf{x}_t', \mathbf{z}_t^{1\prime}\mathbf{C}_0, \mathbf{z}_t'\mathbf{C}\right). \tag{22}$$

From (11), (17), (21) and (10),

$$\boldsymbol{\gamma} \sim N\left(\overline{\boldsymbol{\gamma}}, \overline{\mathbf{H}}_\gamma^{-1}\right), \quad \overline{\mathbf{H}}_\gamma = \underline{\mathbf{H}}_\gamma + h \sum_{t=1}^T \delta_t \mathbf{w}_t \mathbf{w}_t'$$

where

$$\underline{\mathbf{H}}_{\gamma} = \begin{bmatrix} \underline{\mathbf{H}}_{\beta} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{H}}_{\phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \underline{\mathbf{H}}_{\psi} \end{bmatrix},$$

with

$$\underline{\mathbf{H}}_{\phi} = \underline{h}_{\phi} h \mathbf{I}_{m_1 - 1} \text{and } \underline{\mathbf{H}}_{\psi} = \underline{h}_{\psi} h \, \mathrm{Diag}\left(h_1, \ldots, h_{m_1}\right);$$

the mean is $\overline{\boldsymbol{\gamma}} = \overline{\mathbf{H}}_{\gamma}^{-1} \overline{\mathbf{c}}_{\gamma}$ with

$$\overline{\mathbf{c}}_{\gamma} = \underline{\mathbf{c}}_{\gamma} + h \sum_{t=1}^{T} \mathbf{w}_t y_t \delta_t, \;\; \underline{\mathbf{c}}'_{\gamma} = \left(\underline{\boldsymbol{\beta}}' \underline{\mathbf{H}}'_{\beta}, \mathbf{0}'\right).$$

Drawing the state matrix $\mathbf{S}$. The final step of the MCMC algorithm is the draw of the $T \times 2$ matrix of latent states from its distribution conditional on the parameters $\boldsymbol{\theta}$ and observed $\mathbf{X}$ and $\mathbf{Y}$. Define

$$\begin{aligned} d_{tij} &= p\left[y_t \mid \mathbf{s}_t = (i,j), \mathbf{x}_t, \boldsymbol{\theta}\right] \\ &= (2\pi)^{--1/2} \left(h h_i h_{ij}\right)^{1/2} \exp\left[-h h_i h_{ij} \left(y_t - \boldsymbol{\beta}' \mathbf{x}_t - \phi_i - \psi_{ij}\right)^2 \; / \; 2\right] \end{aligned}$$

and

$$d_{ti} = p\left(y_t \mid s_{t1} = i, \mathbf{x}_t, \boldsymbol{\theta}\right) = \sum_{j=1}^{m_2} \rho_{ij} d_{tij}.$$

We draw $\mathbf{s} \sim P\left(\mathbf{s} \mid \mathbf{X}, \mathbf{y}, \boldsymbol{\theta}\right)$ as a two step marginal-conditional, $\mathbf{s}^1 \sim P\left(\mathbf{s}^1 \mid \mathbf{X}, \mathbf{y}, \boldsymbol{\theta}\right)$ followed by $\mathbf{s}^2 \sim P\left(\mathbf{s}^2 \mid \mathbf{s}^1, \mathbf{X}, \mathbf{y}, \boldsymbol{\theta}\right)$. First, given d_{ti} $(t = 1, \ldots, T, i = 1, \ldots, m_1)$ and $\mathbf{P}$, the algorithm of Chib (1996) draws $\mathbf{s}^1 \sim P\left(\mathbf{s}^1 \mid \mathbf{X}, \mathbf{y}, \boldsymbol{\theta}\right)$ and provides $p\left(\mathbf{y} \mid \boldsymbol{\theta}\right)$ as a byproduct of the computations. Then the transitory states s_{t2} are conditionally independent with $P\left(s_{t2} = j \mid s_{t1} = i, y_t, \mathbf{x}_t, \boldsymbol{\theta}\right) \propto \rho_{ij} d_{tij}$.

References

Chib S. 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* **75**: 79-97.

Rao CR. 1965. *Linear Statistical Inference and Its Applications.* New York: Wiley.

Rydén T, Teräsvirta T, Åsbrink S. 1998. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics* **13**: 217-244.