Although the posterior in equation (18) does not belong to any standard parametric family, simulating posterior draws is straightforward. We use a simple auxiliary variable method which inverts the usual gamma-to-Dirichlet scaling transformation used to sample Dirichlet random variables. We explicitly draw $\pi_j^{(i)}$, the normalized transition probabilities out of state j, as

$$\pi_j^{(i)} | \mathbf{z}^{(i)} \sim \mathrm{Dir}\big([\dots, \gamma + n_{jk}^{(i)} + \kappa\delta(j,k), \dots] \odot \mathbf{f}_i\big). \tag{19}$$

The unnormalized transition parameters $\eta_j^{(i)}$ are then given by the deterministic transformation $\eta_j^{(i)} = C_j^{(i)} \pi_j^{(i)}$, where

$$C_j^{(i)} \sim \mathrm{Gamma}\big(K_+^{(i)}\gamma + \kappa, 1\big). \tag{20}$$

Here, $K_+^{(i)} = \sum_k f_{ik}$. This sampling process ensures that transition weights $\boldsymbol{\eta}^{(i)}$ have magnitude entirely informed by the prior, while only the relative proportions are influenced by $\mathbf{z}^{(i)}$. Note that this is a correction to the posterior for $\eta_j^{(i)}$ presented in the earlier work of Fox et al. (2009).

Emission parameters θ_k. The emission parameters $\theta_k = \{\mathbf{A}_k, \Sigma_k\}$ for each feature k have the conjugate matrix normal inverse-Wishart (MNIW) prior of equation (13). Given $\mathbf{z}$, we form θ_k's MNIW posterior using sufficient statistics from observations assigned to state k across all sequences i and time steps t. Letting $\mathbf{Y}_k = \{\mathbf{y}_t^{(i)} : z_t^{(i)} = k\}$ and $\tilde{\mathbf{Y}}_k = \{\tilde{\mathbf{y}}_t^{(i)} : z_t^{(i)} = k\}$, define

$$\begin{aligned} S_{\tilde{y}\tilde{y}}^{(k)} &= \sum_{(t,i)|z_t^{(i)}=k} \tilde{\mathbf{y}}_t^{(i)} \tilde{\mathbf{y}}_t^{(i)^T} + \mathbf{L}, \qquad & S_{y\tilde{y}}^{(k)} &= \sum_{(t,i)|z_t^{(i)}=k} \mathbf{y}_t^{(i)} \tilde{\mathbf{y}}_t^{(i)^T} + \mathbf{ML}, \\ S_{yy}^{(k)} &= \sum_{(t,i)|z_t^{(i)}=k} \mathbf{y}_t^{(i)} \mathbf{y}_t^{(i)^T} + \mathbf{MLM}^T, \qquad & S_{y|\tilde{y}}^{(k)} &= S_{yy}^{(k)} - S_{y\tilde{y}}^{(k)} S_{\tilde{y}\tilde{y}}^{-(k)} S_{y\tilde{y}}^{(k)^T}. \end{aligned} \tag{21}$$

Using standard MNIW conjugacy results, the posterior is then

$$\begin{aligned} \mathbf{A}_k | \Sigma_k, \mathbf{Y}_k, \tilde{\mathbf{Y}}_k &\sim \mathcal{MN}\big(\mathbf{A}_k; S_{y\tilde{y}}^{(k)} S_{\tilde{y}\tilde{y}}^{-(k)}, \Sigma_k, S_{\tilde{y}\tilde{y}}^{(k)}\big), \\ \Sigma_k | \mathbf{Y}_k, \tilde{\mathbf{Y}}_k &\sim \mathrm{IW}\big(|\mathbf{Y}_k| + n_0, S_{y|\tilde{y}}^{(k)} + S_0\big). \end{aligned} \tag{22}$$

Through sharing across multiple time series, we improve inferences about $\{\mathbf{A}_k, \Sigma_k\}$ compared to endowing each sequence with separate behaviors.

6.5. *Sampling the BP and transition hyperparameters.* We additionally place priors on the transition hyperparameters γ and κ, as well as the BP parameters α and c, and infer these via MCMC. Detailed descriptions of these sampling steps are provided in Supplement G.2 of Fox et al. (2014).