equation (12) is useful in reminding the reader that the indices of $\tilde{\pi}_j^{(i)}$ defined by equation (11) are not over 1 to K_i, but rather over the K_i values of k such that $f_{ik} = 1$. Additionally, this notation is useful for concise representations of the posterior distribution.

We construct the model using the unnormalized transition weights $\boldsymbol{\eta}^{(i)}$ instead of just the proper distributions $\boldsymbol{\pi}^{(i)}$ so that we may consider adding or removing states when sampling from the nonparametric posterior. Working with $\boldsymbol{\eta}^{(i)}$ here simplifies expressions, since we need not worry about the normalization constraint required with $\boldsymbol{\pi}^{(i)}$.

4.3. *VAR parameters*. To complete the Bayesian model specification, a conjugate matrix-normal inverse-Wishart (MNIW) prior [cf., West and Harrison (1997)] is placed on the shared collection of dynamic parameters $\theta_k = \{\mathbf{A}_k, \Sigma_k\}$. Specifically, this prior is comprised of an inverse Wishart prior on Σ_k and (conditionally) a matrix normal prior on $\mathbf{A}_k$:

$$\begin{aligned} \Sigma_k | n_0, S_0 &\sim \mathrm{IW}(n_0, S_0), \\ \mathbf{A}_k | \Sigma_k, M, L &\sim \mathcal{MN}(\mathbf{A}_k; M, \Sigma_k, L), \end{aligned} \tag{13}$$

with n_0 the degrees of freedom, S_0 the scale matrix, M the mean dynamic matrix, and L a matrix that together with Σ_k defines the covariance of A_k. This prior defines the base measure B_0 up to the total mass parameter α, which has to be separately assigned (see Section 6.5). The MNIW density function is provided in the supplemental article [Fox et al. (2014)].

5. Model overview. Our beta-process-based featural model couples the dynamic behaviors exhibited by different time series. We term the resulting model the *BP-autoregressive-HMM* (BP-AR-HMM). Figure 3 provides a graphical model representation. Considering the *feature space* (i.e., set of autoregressive parameters) and the *temporal dynamics* (i.e., set of transition distributions) as separate dimensions, one can think of the BP-AR-HMM as a spatio-temporal process comprised of a (continuous) beta process in space and discrete-time Markovian dynamics in time. The overall model specification is summarized as follows:

(1) Draw beta process realization $B \sim \mathrm{BP}(c, B_0)$:

$$B = \sum_{k=1}^{\infty} \omega_k \theta_k \qquad \text{where } \theta_k = \{\mathbf{A}_k, \Sigma_k\}.$$

(2) For each sequence i from 1 to N:
   (a) Draw feature vector $\mathbf{f}_i | B \sim \mathrm{BeP}(B)$.
   (b) Draw feature-constrained transition distributions

$$\pi_j^{(i)} | \mathbf{f}_i \sim \mathrm{Dir}([\ldots, \gamma + \delta(j,k)\kappa, \ldots] \odot \mathbf{f}_i).$$