given the current data assigned to behavior k in $\mathbf{z}$ across all sequences. For new states $k^* \in \{k_a, k_b\}$, we initialize $\hat{\boldsymbol{\theta}}_{k^*}$ from the anchor sequences and then update to account for new data assigned to k^* after each item ℓ. As before, $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}$ are deterministic functions of the conditioning set used to define the *collapsed* proposals for $\mathbf{F}^*, \mathbf{z}^*$; they are discarded prior to subsequent sampling stages.

Merge. To merge k_a, k_b into a new feature k_m, constructing $\mathbf{F}^*$ is deterministic: we set $f^*_{\ell k_m} = 1$ for $\ell \in \mathcal{S}$, and 0 otherwise. We thus need only to sample $\mathbf{z}^*_\ell$ for items in $\mathcal{S}$. We use a block sampler that conditions on $\mathbf{f}^*_\ell, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}^{(\ell)}$, where again $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}^{(\ell)}$ are auxiliary variables.

Accept–reject. After drawing a candidate configuration $(\mathbf{F}^*, \mathbf{z}^*)$, the final step is to compute a Metropolis–Hastings acceptance ratio ρ. Equation (27) gives the ratio for a *split* move which creates features k_a, k_b from k_m:

$$(27)\qquad \rho_{\text{split}} = \frac{p(\mathbf{y}, \mathbf{F}^*, \mathbf{z}^*)}{p(\mathbf{y}, \mathbf{F}, \mathbf{z})} \frac{q_{\text{merge}}(\mathbf{F}, \mathbf{z} | \mathbf{y}, \mathbf{F}^*, \mathbf{z}^*, k_a, k_b)}{q_{\text{split}}(\mathbf{F}^*, \mathbf{z}^* | \mathbf{y}, \mathbf{F}, \mathbf{z}, k_m)} \frac{q_k(k_a, k_b | \mathbf{y}, \mathbf{F}^*, \mathbf{z}^*, i, j)}{q_k(k_m, k_m | \mathbf{y}, \mathbf{F}, \mathbf{z}, i, j)}.$$

Recall that our sampler only updates discrete variables $\mathbf{F}, \mathbf{z}$ and marginalizes out continuous HMM parameters $\boldsymbol{\eta}, \boldsymbol{\theta}$. Our split-merge moves are therefore only tractable with conjugate emission models such as the VAR likelihood and MNIW prior. Proposals which instantiate emission parameters $\boldsymbol{\theta}$, as in Jain and Neal (2007), would be required in the nonconjugate case.

For complete split-merge algorithmic details, consult Supplement F of Fox et al. (2014). In particular, we emphasize that the nonuniform choice of features to split or merge requires some careful accounting, as does the correct computation of the reverse move probabilities. These issues are discussed in the supplemental article [Fox et al. (2014)].

7.3. *Annealing MCMC proposals*. We have presented two novel MCMC moves for adding or deleting features in the BP-AR-HMM: split-merge and birth–death moves. Both propose a new discrete variable configuration $\Psi^* = (\mathbf{F}^*, \mathbf{z}^*)$ with either one more or one fewer feature. This proposal is accepted or rejected with probability $\min(1, \rho)$, where ρ has the generic form

$$(28)\qquad \rho = \frac{p(\mathbf{y}, \Psi^*)}{p(\mathbf{y}, \Psi)} \frac{q(\Psi | \Psi^*, \mathbf{y})}{q(\Psi^* | \Psi, \mathbf{y})}.$$

This Metropolis–Hastings ratio ρ accounts for improvement in joint probability [via the ratio of $p(\cdot)$ terms] and the requirement of reversibility [via the ratio of $q(\cdot)$ terms]. We call this latter ratio the *Hastings factor*. Reversibility ensures that detailed balance is satisfied, which is a sufficient condition for convergence to the true posterior distribution.