

*Selecting anchors.* Following Dahl (2005), we first select distinct anchor data items  $i$  and  $j$  uniformly at random from all time series. The fixed choice of  $i, j$  defines a split-merge transition kernel satisfying detailed balance [Tierney (1994)]. Next, we select from each anchor one feature it possesses, denoted  $k_i, k_j$ , respectively. This choice determines the proposed move: we merge  $k_i, k_j$  if they are distinct, and split  $k_i = k_j$  into two new features otherwise.

Selecting  $k_i, k_j$  uniformly at random is problematic. First, in data sets with many features choosing  $k_i = k_j$  is unlikely, making split moves rare. We need to bias the selection process to consider splits more often. Second, in a reasonably fit model most feature pairs will not make a sensible merge. Selecting a pair that explains similar data is crucial for efficiency. We thus develop a proposal distribution which first draws  $k_i$  uniformly from the positive entries in  $\mathbf{f}_i$ , and then selects  $k_j$  given fixed  $k_i$  as follows:

$$(24) \quad q_k(k_i, k_j | \mathbf{f}_i, \mathbf{f}_j) = \text{Unif}(k_i | \{k : f_{ik} = 1\}) q(k_j | k_i, \mathbf{f}_j),$$

$$(25) \quad q(k_j = k | k_i, \mathbf{f}_j) \propto \begin{cases} 2R_j f_{jk}, & \text{if } k = k_i, \\ f_{jk} \frac{m(\mathbf{Y}_{k_i}, \mathbf{Y}_k)}{m(\mathbf{Y}_{k_i})m(\mathbf{Y}_k)}, & \text{otherwise,} \end{cases}$$

where  $\mathbf{Y}_k$  denotes all observed data in any segment assigned to  $k$  (determined by  $\mathbf{z}$ ) and  $m(\cdot)$  denotes the *marginal likelihood* of pooled data observations under the emission distribution. A high value for the ratio  $\frac{m(\mathbf{Y}_{k_i}, \mathbf{Y}_k)}{m(\mathbf{Y}_{k_i})m(\mathbf{Y}_k)}$  indicates that the model prefers to explain all data assigned to  $k_i, k_j$  together rather than use a separate feature for each. This choice biases selection toward promising merge candidates, leading to higher acceptance rates. We set  $R_j = \sum_{k_j \neq k_i} f_{jk_j} \frac{m(\mathbf{Y}_{k_i}, \mathbf{Y}_{k_j})}{m(\mathbf{Y}_{k_i})m(\mathbf{Y}_{k_j})}$  to ensure the probability of a split (when possible) is  $2/3$ .

For the VAR likelihood of interest, the marginal likelihood  $m(\mathbf{Y}_k)$  of all data assigned to feature  $k$ , integrating over parameters  $\theta_k = \{\mathbf{A}_k, \Sigma_k\}$ , is

$$(26) \quad \begin{aligned} m(\mathbf{Y}_k) &= p(\mathbf{Y}_k | M, L, S_0, n_0) \\ &= \iint p(\mathbf{Y}_k | \mathbf{A}_k, \Sigma_k) p(\mathbf{A}_k | M, \Sigma_k, L) p(\Sigma_k | n_0, S_0) d\Sigma_k d\mathbf{A}_k \\ &= \frac{1}{(2\pi)^{(n_k d)/2}} \cdot \frac{\Gamma_d((n_k + n_0)/2)}{\Gamma_d(n_0/2)} \cdot \frac{|S_0|^{n_0/2}}{|S_{y|\bar{y}}^{(k)}|^{(n_k + n_0)/2}} \cdot \frac{|L|^{1/2}}{|S_{\bar{y}\bar{y}}^{(k)}|^{1/2}}, \end{aligned}$$

where  $\Gamma_d(\cdot)$  is the  $d$ -dimensional multivariate gamma function,  $|\cdot|$  denotes the determinant,  $n_k$  counts the number of observations in set  $\mathbf{Y}_k$ , and sufficient statistics  $S_{\cdot,\cdot}^{(k)}$  are defined in equation (21). Further details on this feature selection process are given in Supplement F.1, especially Algorithm F.2, of Fox et al. (2014).

Once  $k_i, k_j$  are fixed, we construct the candidate state  $\mathbf{F}^*, \mathbf{z}^*$  for the proposed move. This construction depends on whether a split or merge occurs, as detailed