in Figure 6(j). In this plot, we present a histogram of the effective truncation level, L_{eff}, used over the 30,000 Gibbs iterations on three chains. We computed this effective truncation level by summing over the number of state transitions considered during a full sweep of sampling $z_{1:T}$ and then dividing this number by the length of the data set, T, and taking the square root. Finally, on a more technical note, our fixed, truncated model allows for more vectorization of the code than the beam sampler. Thus, in practice, the difference in computation time between the samplers is significantly less than the $O(L^2/L_{\text{eff}}^2)$ factor obtained by counting state transitions.

From this point onward, we present results only from blocked sampling since we have seen the clear advantages of this method over the sequential, direct assignment sampler.

Fast state-switching. In order to warrant the general use of the sticky model, one would like to know that the sticky parameter incorporated in the model does not preclude learning models with fast dynamics. To this end, we explored the performance of the sticky HDP-HMM on data generated from a model with a high probability of switching between states. Specifically, we generated observations from a four-state HMM with the following transition probability matrix:

$$\begin{bmatrix} 0.4 & 0.4 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.4 & 0.4 \\ 0.1 & 0.1 & 0.4 & 0.4 \end{bmatrix}. \tag{6.1}$$

We once again used a truncation level $L = 20$. Since we are restricting ourselves to the blocked Gibbs sampler, it is no longer necessary to use a conjugate base measure. Instead we placed an independent Gaussian prior on the mean parameter and an inverse-Wishart prior on the variance parameter. For the Gaussian prior, we set the mean and variance hyperparameters to be equal to the empirical mean and variance of the entire data set. The inverse-Wishart hyperparameters were set such that the expected variance is equal to 0.75 times that of the entire data set, with three degrees of freedom.

The results depicted in Figure 7 confirm that by inferring a small probability of self-transition, the sticky HDP-HMM is indeed able to capture fast HMM dynamics, and just as quickly as the original HDP-HMM (although with higher variability). Specifically, we see that the histogram of the self-transition proportion parameter ρ for this data set [see Figure 7(d)] is centered around a value close to the true probability of self-transition, which is substantially lower than the mean value of this parameter on the data with high persistence [Figure 6(c)].

6.2. *Multinomial emissions*. The difference in modeling power, rather than simply burn-in rate, between the sticky and original HDP-HMM is more pronounced when we consider multinomial emissions. This is because the multinomial observations are embedded in a discrete topological space in which there