

# **STRATHCLYDE**

**DISCUSSION PAPERS IN ECONOMICS**



## **MODELING U.S. INFLATION DYNAMICS: A BAYESIAN NONPARAMETRIC APPROACH**

**BY**

**MARKUS JOCHMANN**

**No. 10-01**

**DEPARTMENT OF ECONOMICS  
UNIVERSITY OF STRATHCLYDE  
GLASGOW**

# Modeling U.S. Inflation Dynamics: A Bayesian Nonparametric Approach

Markus Jochmann

Department of Economics, University of Strathclyde

markus.jochmann@strath.ac.uk

January 2010

## **Abstract**

This paper uses an *infinite hidden Markov model* (IHMM) to analyze U.S. inflation dynamics with a particular focus on the persistence of inflation. The IHMM is a Bayesian nonparametric approach to modeling structural breaks. It allows for an unknown number of breakpoints and is a flexible and attractive alternative to existing methods. We found a clear structural break during the recent financial crisis. Prior to that, inflation persistence was high and fairly constant.

JEL classifications: C11, C22, E31

Keywords: inflation dynamics, hierarchical Dirichlet process, IHMM, structural breaks, Bayesian nonparametrics

# 1 Introduction

There is ample evidence in the literature that many macroeconomic and financial time series display structural instability (see, e.g., Stock and Watson, 1996, or Ang and Bekaert, 2002). Ignoring this feature in model specification can lead to misleading conclusions and is a main source of poor forecasts. These implications have been shown by, among others, Clements and Hendry (1999) and Koop and Potter (2001).

Possible changes in the inflation process and its persistence have received especially much attention in the literature. Inflation persistence, i.e. the speed with which inflation returns to its base level after a shock, is important for many aspects of macroeconomics in general and monetary policy in particular. Probably most importantly, it is at the heart of the revisionism debate initiated by Taylor (1998). He warned that the decline in the persistence of inflation might lead policymakers to return to the belief that there is an exploitable trade-off between inflation and unemployment in the long run. Additionally, empirical evidence on inflation persistence informs theoretical researchers as to the importance, or lack thereof, of allowing for a dynamically changing inflation persistence in models of price adjustment. Finally, empirical results also help answer the question whether not only monetary policy has changed in the U.S., but also the response of inflation to monetary shocks.

However, the empirical evidence on the properties of inflation persistence in the literature is ambiguous. On one hand, Cogley and Sargent (2001) use a multivariate time-varying parameter model and find that inflation persistence increased in the early 1970s, remained high for around a decade and declined afterwards. Their result is in accordance with the findings of Brainard and Perry (2000) and Taylor (2000). On the other hand, Stock (2001) applies univariate methods and finds that inflation persistence was roughly constant and high over the past 40 years. This view is also supported by Pivetta and Reis (2007).

This article contributes further evidence to this ongoing debate by applying non-parametric Bayesian techniques to model U.S. inflation dynamics. More specifically, we use an *infinite hidden Markov model* (IHMM). The IHMM was introduced by Beal et al. (2002) and Teh et al. (2006) and has been successfully applied to inferential problems in fields like genetics (e.g. Beal and Krishnamurthy, 2006) or visual scene recognition (e.g. Kivinen et al., 2007). However, to our knowledge, the IHMM has not been applied in the econometric literature so far.

The IHMM is a nonparametric Bayesian extension of the *hidden Markov model* (HMM). A nonparametric Bayesian model is a probability model with infinitely many parameters (Bernardo and Smith, 1994), or, in other words, a parametrized model that allows the number of parameters to grow with the number of observations. However, for a given sample size it will only select a finite subset of the available parameters to explain the observations. This means that, unlike the HMM, the IHMM does not fix the number of underlying states a priori, but infers them from the data. Thus, the IHMM is an attractive alternative to existing change-point models that typically either assume a small number of change-points (e.g. Chib, 1998) or assume that the parameters change at each point in time. The latter is referred to as the time-varying parameter (TVP) model (e.g. Cogley and Sargent, 2001). Other approaches that allow for a random number of change-points are Koop and Potter (2007), who propose a model where regime durations have a Poisson distribution, or Giordani et al. (2007), who present a state-space model that accounts for parameter instability and outliers, but does not force the parameters to change at each point in time.

The rest of this paper is organized as follows. In Section 2 we first summarize the *Dirichlet process* and the *hierarchical Dirichlet process*, which are the building blocks of the IHMM. We then discuss the IHMM and an augmented version of it. Finally, we analyze the choice of hyperparameters and prior distributions and point out how inference can be done using Markov chain Monte Carlo methods. Further details on the sampling algorithm are given in the Appendix. Section 3 uses the IHMM to model U.S. inflation dynamics and Section 4 concludes.

## 2 The Infinite Hidden Markov Model

### The Dirichlet Process

The *Dirichlet process* (DP) introduced by Ferguson (1973) is a measure on measures defined by the following property: A random probability measure  $G$  is generated by a DP if for any partition  $B_1, \dots, B_m$  on the space of support of  $G_0$  the vector of probabilities  $[G(B_1), \dots, G(B_m)]$  follows a Dirichlet distribution with parameter vector  $[\alpha G_0(B_1), \dots, \alpha G_0(B_m)]$ . We write  $G \sim \text{DP}(\alpha, G_0)$ , where  $\alpha$  is a positive *precision parameter* and  $G_0$  is a *base measure* defining the expectation,  $E(G) = G_0$ .

Sethuraman (1994) showed that any draw  $G \sim \text{DP}(\alpha, G_0)$  can be represented as

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*}, \quad (1)$$

where  $\{\theta_k^*\}_{k=1}^{\infty}$  represent a set of support points drawn i.i.d. from  $G_0$  and  $\delta_{\theta_k^*}$  is a probability measure concentrated at  $\theta_k^*$ . The probability weights  $\boldsymbol{\pi} = \{\pi_k\}_{k=1}^{\infty}$  are coming from a *stick-breaking process*:

$$\pi_k = \xi_k \prod_{l=1}^{k-1} (1 - \xi_l) \quad \text{with } \xi_l \stackrel{iid}{\sim} \text{Beta}(1, \alpha), \quad (2)$$

which we denote by  $\boldsymbol{\pi} \sim \text{Stick}(\alpha)$ .<sup>1</sup> We can see that any draw  $G$  from a  $\text{DP}(\alpha, G_0)$  is discrete and can be represented as an infinite mixture of point masses  $\delta_{\theta_k^*}$ .

Another representation of the DP that highlights its discrete nature is the Pólya urn scheme of Blackwell and MacQueen (1973). The Pólya urn scheme does not consider  $G$  directly but refers to draws  $\theta_1, \theta_2, \dots$  from  $G$ . Blackwell and MacQueen (1973) show that the conditional distribution of  $\theta_i$  given  $\theta_1, \dots, \theta_{i-1}$  has the following form:

$$\theta_i | \theta_1, \dots, \theta_{i-1} \sim \sum_{j=1}^{i-1} \frac{1}{i-1+\alpha} \delta_{\theta_j} + \frac{\alpha}{i-1+\alpha} G_0. \quad (3)$$

This means that  $\theta_i$  takes on the same value as  $\theta_j$  with probability proportional to 1 and is drawn from the base measure  $G_0$  with probability proportional to  $\alpha$ . Clusters emerge since  $\theta_i$  has a positive probability of being equal to previous draws. Letting  $\theta_1^*, \dots, \theta_K^*$  denote the distinct values taken on by  $\theta_1, \dots, \theta_{i-1}$ , we can express equation (3) as

$$\theta_i | \theta_1, \dots, \theta_{i-1} \sim \sum_{k=1}^K \frac{m_k}{i-1+\alpha} \delta_{\theta_k^*} + \frac{\alpha}{i-1+\alpha} G_0, \quad (4)$$

where  $m_k$  is the number of  $\theta_i$  taking the value  $\theta_k^*$ .

If we further introduce indicator variables  $s_1, s_2, \dots$  with  $s_i = k$  indicating  $\theta_i = \theta_k^*$

---

<sup>1</sup>Another notation for the stick-breaking process is  $\boldsymbol{\pi} \sim \text{GEM}(\alpha)$ , where the letters refer to Griffiths, Engen and McCloskey.

we obtain

$$\Pr(s_i = s | s_1, \dots, s_{i-1}) = \sum_{k=1}^K \frac{m_k}{i-1+\alpha} \delta(s, k) + \frac{\alpha}{i-1+\alpha} \delta(s, K+1), \quad (5)$$

where  $\delta(s, k)$  denotes the Kronecker delta<sup>2</sup>. Equation (5) induces a distribution on partitions and is referred to as the *Chinese restaurant process* (CRP, see Pitman, 2006) which is a helpful metaphor for understanding the properties of the DP. Consider a Chinese restaurant with an unbounded number of tables, each serving a unique dish  $\theta_k^*$ . A new customer  $\theta_i$  entering the restaurant chooses a table  $k$  in proportion to the to number of customers already sitting at that table  $m_k$  and we set  $\theta_i = \theta_k^*$ . With probability proportional to  $\alpha$  he sits at a previously unoccupied table  $K+1$  and we draw  $\theta_{K+1}^* \sim G_0$  and set  $\theta_i = \theta_{K+1}^*$ .

The DP is frequently used as a prior on the parameters in a mixture model which leads to the *Dirichlet process mixture model* (DPM model). Consider a group of observations  $\{x_i\}_{i=1}^N$  with  $x_i \stackrel{\text{ind}}{\sim} F(\theta_i)$ . The parameters  $\{\theta_i\}_{i=1}^N$  are generated from an unknown mixture distribution  $G$  which is drawn from a Dirichlet process  $G \sim \text{DP}(\alpha, G_0)$ . The DPM model can be expressed as follows:

$$\boldsymbol{\pi} \sim \text{Stick}(\alpha), \quad (6)$$

$$s_i \sim \boldsymbol{\pi}, \quad i = 1, \dots, N, \quad (7)$$

$$\theta_k^* \sim G_0, \quad k = 1, \dots, \infty, \quad (8)$$

$$x_i \stackrel{\text{ind}}{\sim} F(\theta_{s_i}^*), \quad i = 1, \dots, N, \quad (9)$$

where  $G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*}$  and  $\theta_i = \theta_{s_i}^*$ . The DPM model is depicted as a graphical model in Figure 1(a).

## The Hierarchical Dirichlet Process

In order to link group-specific DPs, Teh et al. (2006) introduced the *hierarchical Dirichlet process* (HDP).<sup>3</sup> Here, group-specific distributions are conditionally independent given a common base distribution  $G_0$  and follow  $G_j \sim \text{DP}(\alpha, G_0)$ . The common base distribution itself follows a Dirichlet process  $G_0 \sim \text{DP}(\eta, H_0)$ . The

<sup>2</sup>The Kronecker delta is a function of two variables that is 1 if they are equal and 0 otherwise.

<sup>3</sup>For a survey on hierarchical Bayesian nonparametric models see Teh and Jordan (2010).

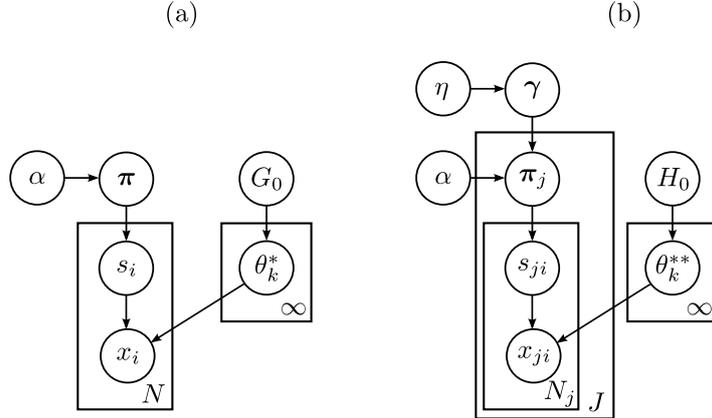


Figure 1: (a) DPM Model and (b) HDPM Model

HDP thus has three parameters: the base measure  $H_0$  and the concentration parameters  $\alpha$  and  $\eta$ . The common base distribution  $G_0$  varies around the prior  $H_0$  where the amount of variability is determined by  $\eta$ . The group-specific distributions  $G_j$  deviate from  $G_0$  with  $\alpha$  governing the amount of variability.

In order to derive a stick-breaking representation for the HDP, we first express the global measure  $G_0$  as:

$$G_0 = \sum_{k=1}^{\infty} \gamma_k \delta_{\theta_k^{**}}, \quad (10)$$

where  $\{\theta_k^{**}\}_{k=1}^{\infty}$  represent a set of support points drawn i.i.d. from  $H_0$  and  $\gamma = \{\gamma_k\}_{k=1}^{\infty} \sim \text{Stick}(\eta)$ . The  $G_j$  reuse the same support points as  $G_0$  but with different proportions:

$$G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\theta_k^{**}}. \quad (11)$$

The weights  $\pi_j = \{\pi_{jk}\}_{k=1}^{\infty}$  are independent given  $\gamma$  (since the  $G_j$  are independent given  $G_0$ ) and one can show that  $\pi_j \stackrel{\text{ind}}{\sim} \text{DP}(\alpha, \gamma)$ .

Teh et al. (2006) also develop a Pólya urn scheme for the HDP and we refer to their paper for technical details on this. The underlying analogue to the CRP is the *Chinese restaurant franchise* (CRF). The CRF consists of  $J$  Chinese restaurants with unboundedly many tables that share a buffet line with unboundedly many dishes. The seating process takes place independently in the restaurants as described before. Then, each table chooses a dish from the franchise-wide buffet line with a probability

proportionally to the number of tables (in the entire franchise) that have previously chosen that dish.

In order to derive the *hierarchical Dirichlet process mixture model* (HDPM model), we consider  $J$  groups of observations  $\{\{x_{ji}\}_{i=1}^{N_j}\}_{j=1}^J$  with  $x_{ji} \stackrel{ind}{\sim} F(\theta_{ji})$ . The parameters  $\{\theta_{ji}\}_{i=1}^{N_j}$  of the  $j$ -th group are generated from an unknown group-specific mixture distribution  $G_j$  for which a HDP prior is assumed. Again, we can consider an indicator variable representation of the HDPM model:

$$\gamma \sim \text{Stick}(\eta) \tag{12}$$

$$\boldsymbol{\pi}_j \stackrel{ind}{\sim} \text{DP}(\alpha, \gamma), \quad j = 1, \dots, J, \tag{13}$$

$$s_{ji} \sim \boldsymbol{\pi}_j, \quad j = 1, \dots, J, \quad i = 1, \dots, N_j, \tag{14}$$

$$\theta_k^{**} \sim H_0, \quad k = 1, \dots, \infty, \tag{15}$$

$$x_{ji} \stackrel{ind}{\sim} F(\theta_{s_{ji}}^{**}), \quad j = 1, \dots, J, \quad i = 1, \dots, N_j, \tag{16}$$

where  $G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\theta_k^{**}}$  and  $\theta_{ji} = \theta_{s_{ji}}^{**}$ . The HDPM model is depicted as a graphical model in Figure 1(b).

## The Infinite Hidden Markov Model

The *infinite hidden Markov model* (IHMM) was introduced by Beal et al. (2002) and Teh et al. (2006). To get from the HDPM model to the IHMM (the IHMM is also referred to as the *hierarchical Dirichlet process hidden Markov model*, HDP-HMM), we start with a finite *hidden Markov model* (HMM). The HMM is a temporal probabilistic model where the state of the underlying process is determined by a single discrete random variable. More formally, we have an unobserved state sequence  $\mathbf{s} = (s_1, \dots, s_T)$  and a sequence of observations  $\mathbf{y} = (y_1, \dots, y_T)$ . Each state variable  $s_t$  can take on a finite number of distinct states:  $1, \dots, K$ . Transitions between the states are Markovian and parametrized by the transition matrix  $\boldsymbol{\pi}$  with  $\pi_{ij} = \Pr(s_t = j | s_{t-1} = i)$ . Each observation  $y_t$  is conditionally independent of the other observations given the state  $s_t$  with the corresponding likelihood depending on a parameter  $\phi_{s_t}$ .

We can write the density of  $y_t$  given the previous state  $s_{t-1}$  as:

$$p(y_t | s_{t-1} = k) = \sum_{s_t=1}^K p(s_t | s_{t-1} = k) p(y_t | s_t) = \sum_{s_t=1}^K \pi_{k,s_t} p(y_t | \phi_{s_t}). \quad (17)$$

We thus have a mixture distribution where the mixture weights  $\boldsymbol{\pi}_k = \{\pi_{k,s_t}\}_{s_t=1}^K$  are specified by  $s_{t-1} = k$  and the mixture component generating  $y_t$  is determined by  $s_t$ . The HMM can thus be interpreted as a set of  $K$  finite mixture models, one for each possible value of  $s_{t-1}$ . Expressed differently, each row of the transition matrix  $\boldsymbol{\pi}$  (indexed by  $s_{t-1}$ ) specifies a different mixture distribution over the same set of mixture components  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)$ .

In order to derive a nonparametric version of the HMM with an unbounded set of states, we replace the finite mixture distributions with Dirichlet process mixtures, again one for each possibly visited state in the previous period. However, we need to couple the Dirichlet process mixtures in such a way that they share the same set of states. This can be done using a HDP mixture and we finally obtain the IHMM:

$$\boldsymbol{\gamma} \sim \text{Stick}(\eta). \quad (18)$$

$$\boldsymbol{\pi}_k \sim \text{DP}(\alpha, \boldsymbol{\gamma}), \quad k = 1, \dots, \infty, \quad (19)$$

$$s_t \sim \text{Multinomial}(\boldsymbol{\pi}_{s_{t-1}}), \quad t = 1, \dots, T, \quad s_0 = 1, \quad (20)$$

$$\phi_k \sim H, \quad k = 1, \dots, \infty, \quad (21)$$

$$y_t \sim F(\phi_{s_t}), \quad t = 1, \dots, T, \quad (22)$$

The IHMM is shown as a graphical model in Figure 2 (for now, we ignore  $\kappa$  which will be introduced in the next section).

### The Sticky IHMM

Equation (19) shows that each row of the transition matrix is drawn from the same DP and, thus, the IHMM does not differentiate between self-transitions and transitions to other states. However, many economic time series exhibit state persistence, and we would like to incorporate this feature into the prior in order to rule out unrealistic high dynamics in the state sequence. Fox et al. (2007, 2008) address this issue by introducing the so called *Sticky IHMM* and we follow their approach in this

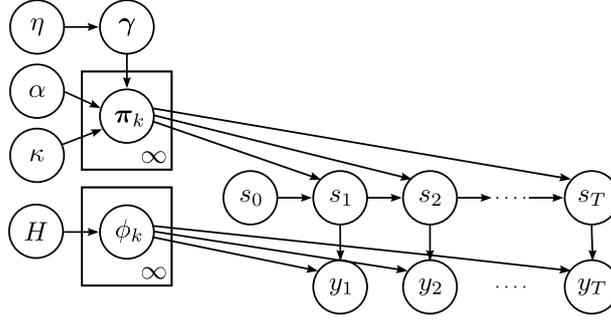


Figure 2: The infinite hidden Markov model (IHMM)

paper. Their idea is to increase the prior probability  $E(\pi_{kk})$  of a self-transition by introducing a positive parameter  $\kappa$  into equation (19) which then becomes:

$$\pi_k | \alpha, \gamma, \kappa \sim \text{DP} \left( \alpha + \kappa, \frac{\alpha \gamma + \kappa \delta_k}{\alpha + \kappa} \right), \quad k = 1, \dots, \infty. \quad (19^*)$$

Thus, an amount  $\kappa$  is added to the  $k$ -th component of  $\alpha \gamma$  which leads to an increased probability of self-transitions. Note that the original IHMM can be obtained by setting  $\kappa = 0$ .

The metaphor that Fox et al. (2007, 2008) develop for their extended model is the CRF with loyal customers. Each restaurant now has a specialty dish that has the same index as the restaurant. This dish is served everywhere (since the restaurants still share the same buffet line) but is more popular in its namesake restaurant. In other words, each restaurant now has a specific rating of the buffet line that puts more weight on the specialty dish.

### Hyperparameters and Prior Distributions

First, we must specify the distribution of the observations  $F(y_t | \phi_{s_t})$  and the base measure  $H$ . In our application to inflation dynamics we assume that the observations are normally distributed:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_{s_t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2), \quad t = 1, \dots, T. \quad (23)$$

We then choose a normal-inverse gamma distribution as base measure:

$$\boldsymbol{\beta}_k | \sigma_k^2 \sim \text{N}(\mathbf{b}_0, \sigma_k^2 \mathbf{B}_0), \quad \sigma_k^2 \sim \text{Inv-Gamma}\left(\frac{c_0}{2}, \frac{d_0}{2}\right), \quad k = 1, \dots, \infty. \quad (24)$$

Note that the normal-inverse gamma distribution is conjugate which leads to straightforward and efficient sampling of the  $\boldsymbol{\beta}_j$  and  $\sigma_j^2$ . However, non-conjugate cases could be handled as well with only minor modifications. We treat the hyperparameters  $\mathbf{b}_0, \mathbf{B}_0, c_0, d_0$  as fixed; another approach would be to place further prior distributions on them.

In contrast, the concentration parameters  $\alpha$  and  $\eta$  and the self-transition parameter  $\kappa$  are treated as unknown quantities which we learn from the data by performing full Bayesian inference. Fox et al. (2007, 2008) show that it is convenient not to work with  $\alpha$  and  $\kappa$  directly but instead with  $\alpha + \kappa$  and  $\rho = \kappa / (\alpha + \kappa)$  and place the following prior distributions on them:

$$\alpha + \kappa \sim \text{Gamma}(e_0, f_0), \quad (25)$$

$$\rho \sim \text{Beta}(g_0, h_0). \quad (26)$$

Finally,  $\eta$  is given a gamma prior:

$$\eta \sim \text{Gamma}(r_0, s_0). \quad (27)$$

### Inference via MCMC Sampling

Since the sticky IHMM is too complex to be analyzed analytically, we need to resort to MCMC sampling techniques (for a comprehensive survey on these methods see, for example, Robert and Casella, 2004). In principle, it is straightforward to set up a Gibbs sampler that alternates between drawing the state sequence, the parameters and the hyperparameters. However, a sampler that sequentially updates each state given all other state assignments generally mixes very slowly due to strong dependencies between consecutive time points.

For this reason, it is more efficient to sample the whole state sequence in one block. However, common dynamic programming algorithms, like the forward-backward algorithm (Rabiner, 1989), cannot be applied because of the infinite number of states.

One solution to this problem is to work with a finite approximation to the DP (Ishwaran and Zarepour, 2002) which is done in Fox et al. (2007, 2008). Another option is to follow Van Gael et al. (2008) who propose *beam sampling* for the IHMM. Their algorithm uses the concept of *slice sampling* (Neal, 2003) and is related to the approach of Walker (2007) for DPM models. The basic idea is to augment the parameter space with a set of auxiliary variables  $\mathbf{u} = (u_1, \dots, u_T)$ . These auxiliary variables do not change the marginal distributions of the other variables but adaptively reduce the set of all valid state sequences to a finite one, such that dynamic programming techniques can be applied.

In our application, we use the beam sampling algorithm for drawing the state sequence. The Gibbs sampling steps for the parameters and hyperparameters are the same as in Fox et al. (2007, 2008). The complete MCMC sampling algorithm is described in the Appendix. For further details and derivations we refer to the original articles.

### 3 U.S. Inflation Dynamics

In this section we employ the sticky IHMM to analyze the dynamics of U.S. inflation. We measure the price level  $P_t$  using seasonally adjusted quarterly data on the PCE deflator obtained from the Bureau of Economic Analysis. Annualized quarterly inflation is then calculated as  $\pi_t = 400 \ln(P_t/P_{t-1})$ . Our sample goes from 1953:I to 2009:III and we use earlier data to initialize the lags of our model.<sup>4</sup>

The inflation series is plotted in Figure 3. Starting out low, inflation rose during the 1970s, reaching a first peak at 11.7% in 1974 and a second peak at 11.8% in 1980. Then, the restrictive monetary policy of the Federal Reserve under Paul Volcker succeeded in lowering inflation to 2.6% in 1983. Afterwards, the inflation rate remained rather stable, with the exception of 2008, when it experienced a sharp drop during the recent banking crisis.

Table 1 includes summary statistics for five different periods of the overall sample. The first-order autocorrelation, which gives us a first indication on the persistence of inflation, was rather low before 1965. During the subsequent 20 years, it was much higher, but declined again after 1985. In the end it was even lower than at the start

---

<sup>4</sup>The starting point of the sample is the same as in Nelson and Schwert (1977) and Stock and Watson (2007).

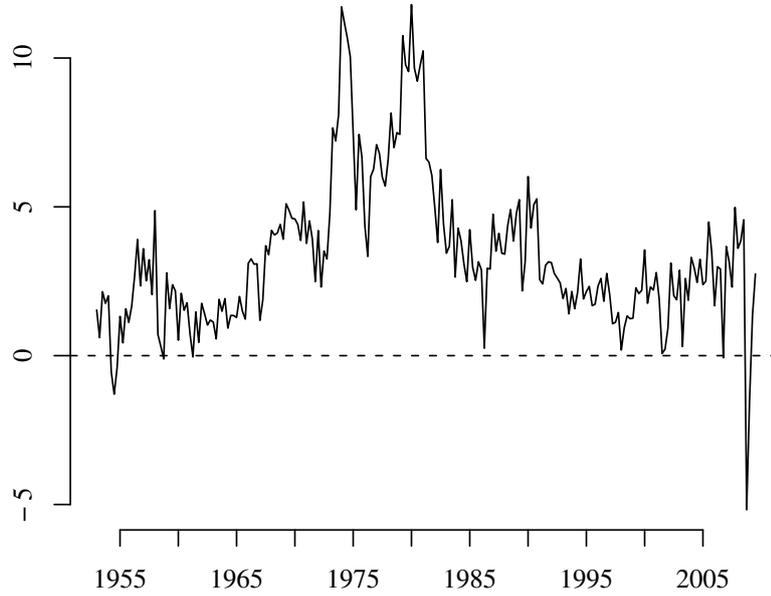


Figure 3: U.S. Inflation Dynamics

of the sample.

As stated above, we assume that inflation is normally distributed and choose to work with a 4th-order autoregressive (AR) representation. Equation (23) becomes

$$\pi_t = \beta_{0,s_t} + \sum_{i=1}^4 \beta_{i,s_t} \pi_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2), \quad t = 1, \dots, T. \quad (23^*)$$

Thus,  $y_t = \pi_t$ ,  $\mathbf{x}_t = (1, \pi_{t-1}, \dots, \pi_{t-4})$  and  $\boldsymbol{\beta}_{s_t} = (\beta_{0,s_t}, \beta_{1,s_t}, \dots, \beta_{4,s_t})$ . We use the

Period	1953:I - 1964:IV	1965:I - 1974:IV	1975:I - 1984:IV	1985:I - 1994:IV	1995:I - 2009:III
Mean	1.489	4.533	6.420	3.223	2.049
S.D.	1.144	2.651	2.380	1.217	1.529
Autocorrelation	0.359	0.825	0.747	0.447	0.250

Table 1: Summary Statistics for Different Periods

prior distributions stated in equations (24) - (27) and choose the prior parameters in the following way. We set  $\mathbf{b}_0 = 0$  and assume  $\mathbf{B}_0$  to be diagonal with the prior variance of the intercept equal to 5 and the prior variances of the AR coefficients equal to 1. Further, we set  $c_0 = 5$  and  $d_0 = 3$ , which implies that  $\sigma_j^2$  has a prior mean of 1.0 and a prior variance of 2.0. The prior distributions for the hyperparameters are assumed to be rather uninformative:  $e_0 = 125$ ,  $f_0 = 5$ ,  $g_0 = 10$ ,  $h_0 = 1$ ,  $r_0 = 5$ ,  $s_0 = 1$ . Our results are based on every 50-th of 500,000 samples from the MCMC output after a burn-in period of 50,000 iterations.<sup>5</sup>

A histogram of the number of inferred states is shown in the top panel of Figure 4. The posterior mode is 5, but we see that the MCMC sampler averages over a large set of values ranging from 2 to 15. The middle two panels of Figure 4 show estimates of the state sequence at two randomly picked iterations of the MCMC sampler. In panel (b), which shows the estimates at iteration 100,000, the sequence consists of 6 different states, in panel (c), which gives the estimates at iteration 200,000, of 4 states. However, not only the numbers of states differ but also the patterns of the sequences. In panel (b) most observations belong to either state 1 or 3, and the sequence switches rather often between these two states. In panel (c), most observations are in state 1 or in state 2, and the sequence switches only once from state 1 to state 2 and once back. This example shows that the data are not overly informative about the actual state pattern. Therefore, it is very important to employ a flexible framework like the IHMM in modeling. The two estimated state sequences also demonstrate the IHMM's capability of dealing with outliers. In panel (b), the observations at 1954Q3 and 2008Q4 are identified as outliers, each being the only observation in the respective state. Similarly, only a few observations occupy states 3 and 4 in panel (c).<sup>6</sup> Finally, the bottom panel of Figure 4 shows posterior means of the break probabilities  $\Pr(s_t \neq s_{t-1})$ . The three peaks are dated 1973Q1, where we have a posterior break probability of 0.458, 1981Q2 with a posterior break probability of 0.567 and 2008Q4 with a posterior break probability of 0.998.

Figure 5 displays posterior means and 10% and 90% quantiles for the intercept, the variance and the sum of the AR coefficients  $\sum_{i=1}^4 \beta_{i,st}$ . The latter serves as

---

<sup>5</sup>The algorithm is coded in C++. It takes around 30 minutes to draw 550,000 samples using a 3 GHz Intel (R) Core (TM) 2 Quad processor (employing a non-parallelized version of the code).

<sup>6</sup>In order to accommodate outliers, we also experimented with a version of the model where the observations were assumed to be drawn from a Student's t-distribution. However, the results did not differ substantially from those based on the normal distribution presented here.

our measure of persistence.<sup>7</sup> The intercept displays some variabilities around 1975 and at the end of the sample, otherwise, it stays rather constant. However, the credible interval is rather wide. The variance is more fluctuating, being highest between 1973Q2 and 1981Q1. However, the credible set is very wide as well, and we cannot rule out a constant variance level. Finally, the sum of the AR coefficients is highest between 1973Q1 and 1974Q1 and between 1976Q4 and 1981Q1. Our measure of persistence displays a clear structural break during the recent banking crisis. Furthermore, the 90% posterior quantile always stays close to 1, and the credible set includes 1 at 38% of the points in the sample period. These results lead to the conclusion that, with the exception of the end of the sample, inflation persistence was high and nearly constant. However, the credible interval is very wide. Therefore, a considerable amount of uncertainty about the exact properties of inflation persistence remains.

Figure 6 presents the outcome of a prior sensitivity analysis focusing on inflation persistence. We argued above that our main results are based on rather uninformative priors for the hyperparameters. In order to verify this, we employed three more informative priors, each of them changing one pair of hyperparameters compared to the prior used in the main analysis. First, we set  $r_0 = 100$  and  $s_0 = 10$ . This forces  $\eta$  to be higher and, thus, leads to a global transition distribution  $\gamma$  that is not as sparse as the original one. The top panel shows posterior means and 10% and 90% quantiles for the sum of the AR coefficients under this prior. Comparing these results with our main results in Figure 5(c), we see that they are nearly the same. The results do not change much either if we force  $\alpha + \kappa$  to be higher by setting  $e_0 = 1000$  and  $f_0 = 25$  (see the panel in the middle). Finally, we set  $g_0 = h_0 = 5$ , which implies a smaller number of self-transitions. The result is shown in the bottom panel. We see that inflation persistence is more bumpy, and the credible intervals are slightly narrower. However, the main conclusions about the properties of inflation persistence do not change.

---

<sup>7</sup>For a discussion of this persistence measure and possible alternatives see Pivetta and Reis (2007). We also calculated the largest autoregressive root as another measure of persistence and obtained results that lead to the same conclusions.

## 4 Conclusions

We applied the infinite hidden Markov model (IHMM) to analyze U.S. inflation dynamics. The IHMM is a Bayesian nonparametric extension of the hidden Markov model (HMM). This means it does not fix the number of states a priori but learns it from the data. Thus, the IHMM is a convenient and flexible approach to model economic time series allowing for an unknown number of structural breaks.

We used the described MCMC algorithm for posterior inference and focused on the sum of AR coefficients as a measure of inflation persistence. We found a clear structural break during the recent financial crisis. Prior to that, inflation persistence was high and approximately constant since 1953. However, the credible intervals were wide; thus, a substantial amount of uncertainty about inflation dynamics remained.

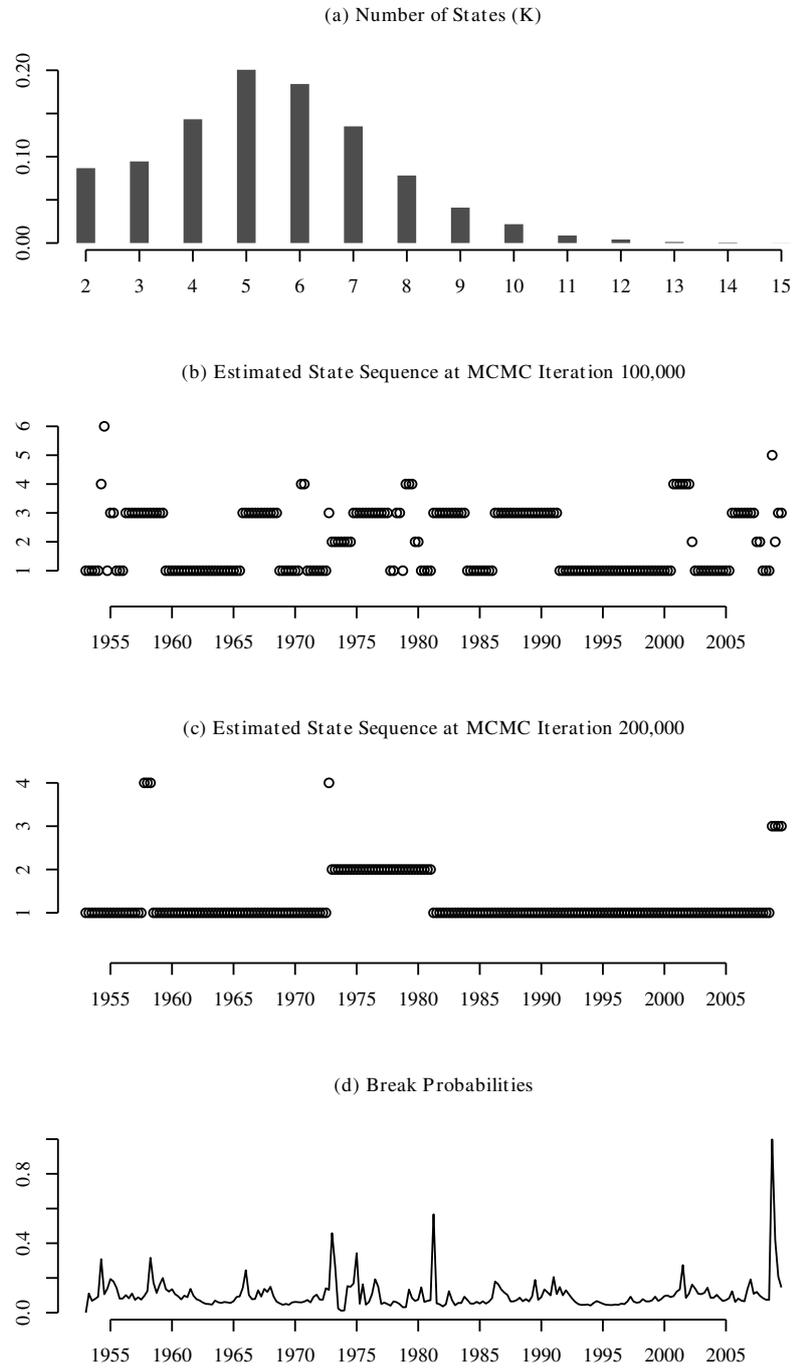


Figure 4: (a) Number of Different States (K), (b) and (c) State Sequence at Two Iterations of the MCMC Sampler, and (d) Posterior Break Probabilities

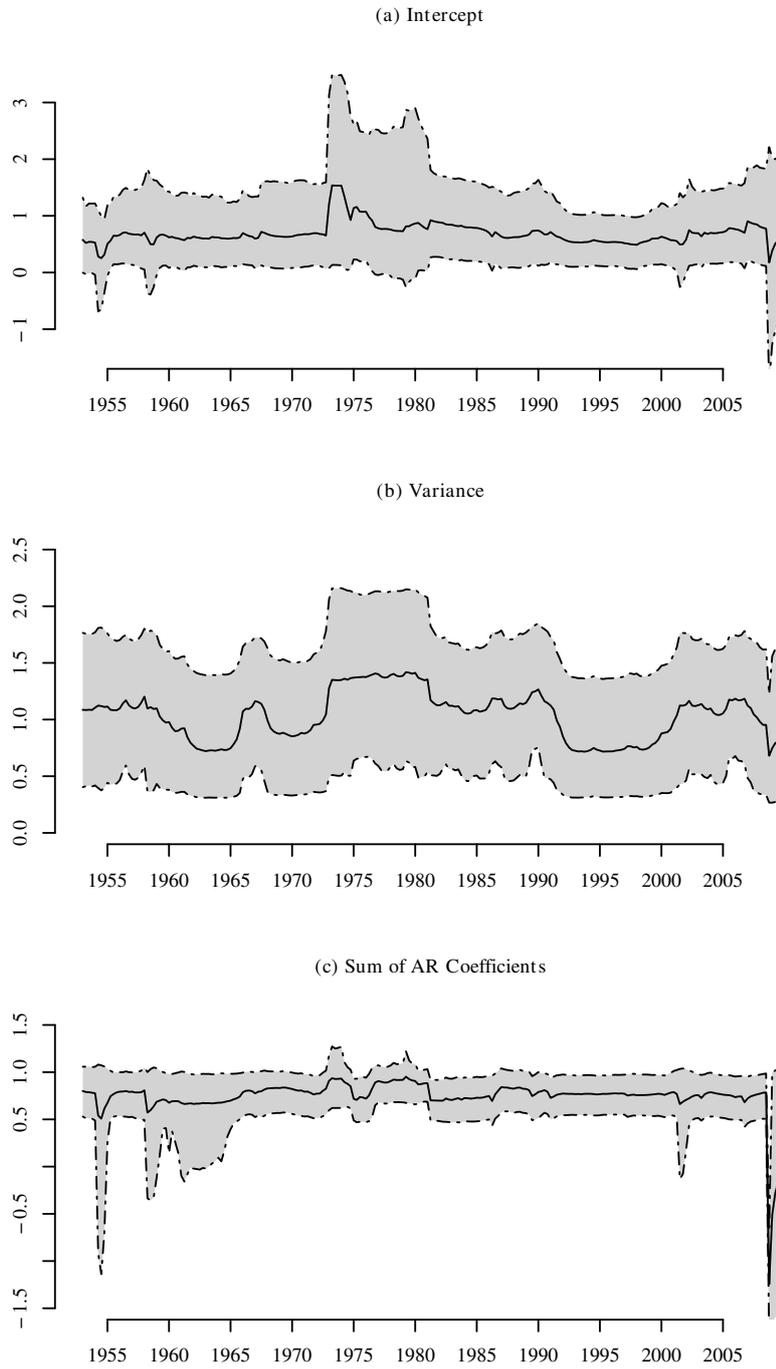
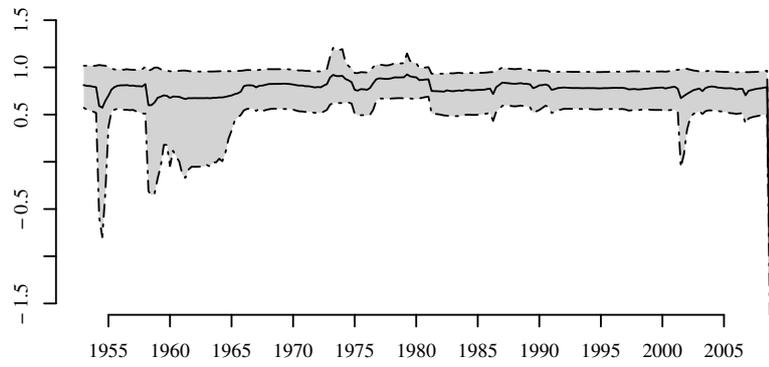
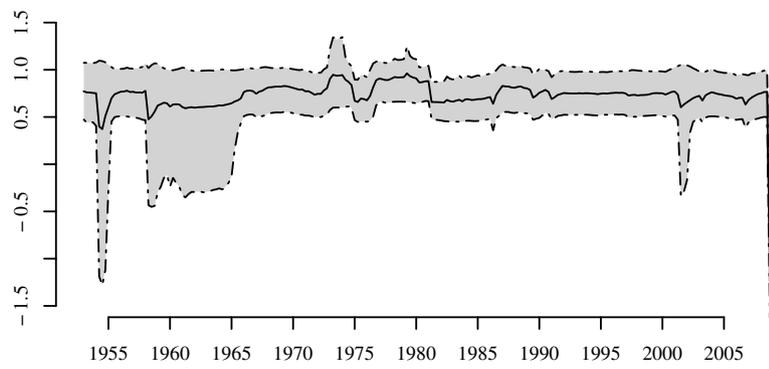


Figure 5: Means and 10% and 90% Quantiles of the Posterior Distributions: (a) Intercept, (b) Variance, and (c) Sum of AR Coefficients

(a) Alternative Prior 1



(b) Alternative Prior 2



(c) Alternative Prior 3

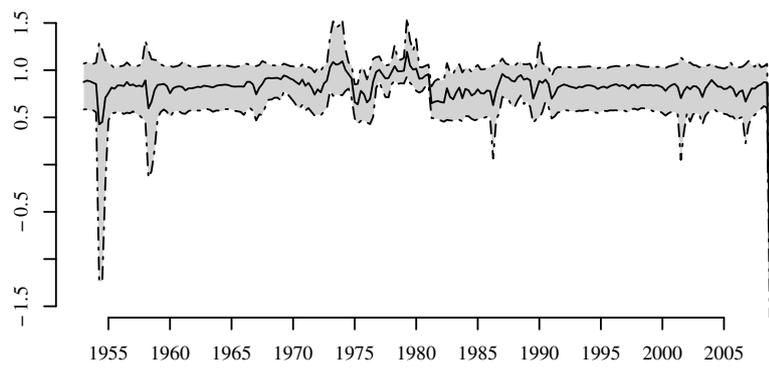


Figure 6: Prior Sensitivity Analysis

## References

- Ang, A., Bekaert, G., 2002. Regime switches in interest rates. *Journal of Business & Economic Statistics*, 20, 163–182.
- Beal, M., Krishnamurthy, P., 2006. Gene expression time course clustering with countably infinite hidden Markov models. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*.
- Beal, M. J., Ghahramani, Z., Rasmussen, C. E., 2002. The infinite hidden Markov model. *Advances in Neural Information Processing Systems* 14, 577–584.
- Bernardo, J. M., Smith, A. F. M., 1994. *Bayesian Theory*. Wiley, New York.
- Blackwell, D., MacQueen, J., 1973. Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* 1, 353–355.
- Brainard, W., Perry, G., 2000. Making policy in a changing world. In: Perry, G., Tobin, J. (Eds.), *Economic Events, Ideas, and Policies: The 1960s and After*. Brookings Institution Press, Washington.
- Chib, S., 1998. Estimation and comparison of multiple change point models. *Journal of Econometrics* 86, 221–241.
- Clements, M. P., Hendry, D. F., 1999. *Forecasting Non-Stationary Economic Time Series*. MIT Press, Cambridge, MA.
- Cogley, T., Sargent, T. J., 2001. Evolving post-world war II U.S. inflation dynamics. *NBER Macroeconomics Annual* 16, 331–373.
- Ferguson, T. S., 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1, 209–230.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., Willsky, A. S., 2007. The sticky HDP-HMM: Bayesian nonparametric hidden Markov models with persistent states. Tech. Rep. 2777, MIT Laboratory for Information and Decision Systems.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., Willsky, A. S., 2008. An HDP-HMM for systems with state persistence. *Proceedings of the 25th International Conference on Machine Learning*.

- Giordani, P., Kohn, R., van Dijk, D., 2007. A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics* 137, 112–133.
- Ishwaran, H., Zarepour, M., 2002. Exact and approximate sum-representations for the Dirichlet process. *The Canadian Journal of Statistics* 30, 269–283.
- Kivinen, J. J., Sudderth, E. B., Jordan, M. I., 2007. Learning multiscale representations of natural scenes using Dirichlet processes. *Proceedings of the IEEE International Conference on Computer Vision*.
- Koop, G., Potter, S., 2001. Are apparent findings of nonlinearity due to structural instability in economic time series? *The Econometrics Journal* 4, 37–55.
- Koop, G., Potter, S., 2007. Forecasting and estimating multiple change-point models. *Review of Economic Studies* 74, 763–789.
- Neal, R. M., 2003. Slice sampling. *Annals of Statistics* 31, 705–741.
- Nelson, C. R., Schwert, G. W., 1977. Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant. *American Economic Review* 67, 478–486.
- Pitman, J., 2006. *Combinatorial Stochastic Processes*. Springer, Berlin.
- Pivetta, F., Reis, R., 2007. The persistence of inflation in the United States. *Journal of Economic Dynamics and Control* 31, 1326–1358.
- Rabiner, L. R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–286.
- Robert, C. P., Casella, G., 2004. *Monte Carlo Statistical Methods*. Springer, New York.
- Sethuraman, J., 1994. A constructive definition of Dirichlet priors. *Statistica Sinica* 4, 639–650.
- Stock, J., Watson, M. W., 1996. Evidence of structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics* 14, 11–30.

- Stock, J. H., 2001. Evolving post-world war II U.S. inflation dynamics: Comment. NBER Macroeconomics Annual 16, 379–387.
- Stock, J. H., Watson, M. W., 2007. Why has U.S. inflation become harder to forecast? Journal of Money, Credit and Banking 39, 3–33.
- Taylor, J., 1998. Monetary policy guidelines for unemployment and inflation stability. In: Friedman, B., Solow, R. (Eds.), Inflation, Unemployment and Monetary Policy. MIT Press, Cambridge.
- Taylor, J., 2000. Low inflation, pass-through, and the pricing power of firms. European Economic Review 44, 1389–1408.
- Teh, Y. W., Jordan, M. I., 2010. Hierarchical Bayesian nonparametric models with applications. In: Hjort, N. L., Holmes, C., Müller, P. (Eds.), Bayesian Nonparametrics. Cambridge University Press, Cambridge.
- Teh, Y. W., Jordan, M. I., Beal, M. J., Blei, D. M., 2006. Hierarchical Dirichlet processes. Journal of the American Statistical Association 101, 1566–1581.
- Van Gael, J., Saatci, Y., Teh, Y. W., Ghahramani, Z., 2008. Beam sampling for the infinite hidden Markov model. Proceedings of the 25th International Conference on Machine Learning.
- Walker, S. G., 2007. Sampling the Dirichlet mixture model with slices. Communications in Statistics 36, 45–54.

## Appendix: Implementation of the MCMC Sampler

This Appendix gives details on the MCMC sampler which combines the beam sampling algorithm of Van Gael et al. (2008) with the sampling techniques for the sticky IHMM derived in Fox et al. (2007, 2008). The parameters that need to be sampled are the hyperparameters  $\eta$ ,  $\alpha$  and  $\kappa$ , the global transition distribution  $\gamma$ , the transition distributions  $\boldsymbol{\pi} = \{\boldsymbol{\pi}_k\}_{k=1}^K$ , the state sequence  $\mathbf{s} = \{s_t\}_{t=1}^T$ , and the parameters of the outcome distributions  $\boldsymbol{\theta} = \{\boldsymbol{\beta}_k, \sigma_k^2\}_{k=1}^K$ .  $K$  denotes the number of distinct states (which are labeled  $1, \dots, K$ ) and changes during sampling. The auxiliary variables  $\mathbf{u} = \{u_t\}_{t=1}^T$  are introduced to make the set of possible state sequences finite. Another set of auxiliary variables consists of  $\mathbf{m} = \{\{m_{jk}\}_{j=1}^K\}_{k=1}^K$ , where (in terms of the Chinese restaurant franchise)  $m_{jk}$  denotes the number of tables in restaurant  $j$  that were served dish  $k$ ,  $\mathbf{r} = \{r_k\}_{k=1}^K$ , where  $r_k$  denotes the number of tables in restaurant  $k$  that eat the namesake dish  $k$  but originally considered to eat another dish (and finally were overridden due to the increased probability of a self-transition), and  $\bar{\mathbf{m}} = \{\{\bar{m}_{jk}\}_{j=1}^K\}_{k=1}^K$ , where  $\bar{m}_{jk}$  denotes the number of tables in restaurant  $j$  that considered to eat dish  $k$ . The auxiliary variables  $\mathbf{v} = \{v_k\}_{k=1}^K$ ,  $\mathbf{w} = \{w_k\}_{k=1}^K$ ,  $\nu$ , and  $\lambda$  are helpful for sampling the hyperparameters. Finally,  $n_{jk}$  counts the number of transitions from  $j$  to  $k$  in the state sequence  $\mathbf{s}$ . Sums are denoted by dots, i.e.  $x_{.b} = \sum_a x_{ab}$ ,  $x_{a.} = \sum_b x_{ab}$ , and  $x_{..} = \sum_b \sum_a x_{ab}$ . The MCMC sampler then consists of the following steps:

- (0) **Initialize parameters:** Choose a starting value for  $K$  and initialize all parameters. The infinitely many states that do not occur in  $\mathbf{s}$  are merged into one state. Thus  $\gamma$  and each  $\boldsymbol{\pi}_k$  have  $K + 1$  elements.
- (1) **Sampling  $\mathbf{u}$ :** For  $t = 1, \dots, T$ , sample  $u_t$  from  $\varphi(u_t | \boldsymbol{\pi}, s_{t-1}, s_t) = \text{U}(0, \pi_{s_{t-1}s_t})$ , where  $\text{U}(a, b)$  denotes the uniform distribution on the interval  $(a, b)$ .
- (2) **Sampling  $\mathbf{s}$ :**
  - (i) **If necessary, break  $\boldsymbol{\pi}$  and  $\gamma$ :** While  $\max(\{\pi_{k,K+1}\}_{k=1}^K) > \min(\{u_t\}_{t=1}^T)$ , repeat the following steps:
    - (a) Draw  $\boldsymbol{\pi}_{K+1} \sim \text{Dirichlet}(\alpha\boldsymbol{\gamma})$ .
    - (b) Break the last element of  $\boldsymbol{\gamma}$ .

- (b-1) Draw  $\zeta \sim \text{Beta}(1, \eta)$ .
  - (b-2) Add element  $\gamma_{K+2} = (1 - \zeta)\gamma_{K+1}$ .
  - (b-3) Set  $\gamma_{K+1} = \zeta\gamma_{K+1}$ .
  - (c) Break the last element of each  $\boldsymbol{\pi}_k$ . For  $k = 1, \dots, K + 1$ :
    - (c-1) Draw  $\zeta_k \sim \text{Beta}(\alpha\gamma_{K+1}, \alpha\gamma_{K+2})$ .
    - (c-2) Add element  $\pi_{k,K+2} = (1 - \zeta_k)\pi_{k,K+1}$ .
    - (c-3) Set  $\pi_{k,K+1} = \zeta_k\pi_{k,K+1}$ .
  - (d) Sample  $\sigma_{K+1}^2 \sim \text{Inv-Gamma}(\frac{c_0}{2}, \frac{d_0}{2})$  and  $\boldsymbol{\beta}_{K+1} \sim \text{N}(\mathbf{b}_0, \sigma_{K+1}^2 \mathbf{B}_0)$ .
  - (e) Increment  $K$ .
- (ii) **Sample  $\mathbf{s}$ :** Sample  $\mathbf{s}$  from  $\varphi(\mathbf{s}|\boldsymbol{\pi}, \mathbf{u}, \boldsymbol{\theta})$ :
- (a) Working sequentially forward in time, calculate
$$m_t(k) = \text{P}(y_1, \dots, y_t | s_t = k, \mathbf{u}):$$
    - (a-1) Initialize  $m_1(k) = 1(u_1 < \pi_{1k})\text{N}(y_1 | x_1' \boldsymbol{\beta}_k, \sigma_k^2)$  for  $k = 1, \dots, K$ .
    - (a-2) Induce  $m_t(k) = \sum_{l=1}^K 1(u_t < \pi_{lk})\text{N}(y_t | x_t' \boldsymbol{\beta}_k, \sigma_k^2) m_{t-1}(l)$  for  $t = 2, \dots, T$  and  $k = 1, \dots, K$ .
  - (b) Working sequentially backwards in time, sample state indicators:
    - (b-1) Sample  $s_T$  from  $\sum_{k=1}^K m_T(k) \delta(s_T, k)$ .
    - (b-2) Sample  $s_t$  from  $\sum_{k=1}^K m_t(k) \delta(s_t, k) 1(u_{t+1} < \pi_{k, s_{t+1}})$  for  $t = T - 1, \dots, 1$ .
- (3) **Cleaning up:** Remove the redundant states and relabel the remaining ones from  $1, \dots, K$ . Adapt  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\pi}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\sigma}^2$  accordingly.
- (4) **Sampling auxiliary variables  $\mathbf{m}$ ,  $\mathbf{r}$  and  $\bar{\mathbf{m}}$ :**
- (i) **Sample  $\mathbf{m}$ :** For  $j = 1, \dots, K$  and  $k = 1, \dots, K$ , sample  $m_{jk}$  from  $\varphi(m_{jk} | \mathbf{s}, \gamma_k, \alpha, \kappa)$  as follows: Set  $m_{jk} = 0$ . For  $i = 1, \dots, n_{jk}$ , sample  $x_i \sim \text{Bernoulli}\left(\frac{\alpha\gamma_k + \kappa\delta(j,k)}{i-1 + \alpha\gamma_k + \kappa\delta(j,k)}\right)$ . If  $x_i = 1$  increment  $m_{jk}$ .
  - (ii) **Sample  $\mathbf{r}$ :** For  $j = 1, \dots, K$ , sample  $r_j$  from  $\varphi(r_j | m_{jj}, \gamma_j, \alpha, \kappa) = \text{Binomial}\left(m_{jj}, \frac{\rho}{\rho + \gamma_j(1-\rho)}\right)$ , where  $\rho = \frac{\kappa}{\alpha + \kappa}$ .
  - (iii) **Update  $\bar{\mathbf{m}}$ :** For  $j = 1, \dots, K$  and  $k = 1, \dots, K$ , set  $\bar{m}_{jk} = m_{jk}$  if  $j \neq k$ , set  $\bar{m}_{jk} = m_{jk} - r_j$  if  $j = k$ .

- (5) **Sampling  $\boldsymbol{\gamma}$ :** Draw  $\boldsymbol{\gamma}$  from  $\varphi(\boldsymbol{\gamma}|\overline{\mathbf{m}}, \eta) = \text{Dirichlet}(\overline{\mathbf{m}}_{\cdot 1}, \dots, \overline{\mathbf{m}}_{\cdot K}, \eta)$ .
- (6) **Sampling  $\boldsymbol{\pi}$ :** For  $k = 1, \dots, K$ , sample  $\boldsymbol{\pi}_k$  from  
 $\varphi(\boldsymbol{\pi}_k|\boldsymbol{\gamma}, \mathbf{s}, \alpha, \kappa) = \text{Dirichlet}(\alpha\gamma_1 + n_{k1}, \dots, \alpha\gamma_k + \kappa + n_{kk}, \dots, \alpha\gamma_K + n_{kK}, \alpha\gamma_{K+1})$ .
- (7) **Sampling  $\boldsymbol{\theta}$ :** For  $k = 1, \dots, K$ , sample  $\sigma_k^2 \sim \text{Inv-Gamma}(\frac{c_*}{2}, \frac{d_*}{2})$  and  $\boldsymbol{\beta}_{K+1} \sim \text{N}(\mathbf{b}_*, \sigma_{K+1}^2 \mathbf{B}_*)$  with  $\mathbf{B}_* = (\mathbf{B}_0^{-1} + \sum_{t:s_t=k} \mathbf{x}_t \mathbf{x}_t')^{-1}$ ,  $\mathbf{b}_* = \mathbf{B}_* (\mathbf{B}_0^{-1} \mathbf{b}_0 + \sum_{t:s_t=k} \mathbf{x}_t y_t)$ ,  $c_* = c_0 + \sum_{t:s_t=k} 1$  and  $d_* = d_0 + s^2 + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*)' (\sum_{t:s_t=k} \mathbf{x}_t \mathbf{x}_t') \mathbf{B}_* \mathbf{B}_0^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*)$ , where  $\hat{\boldsymbol{\beta}} = (\sum_{t:s_t=k} \mathbf{x}_t \mathbf{x}_t')^{-1} \sum_{t:s_t=k} \mathbf{x}_t y_t$  and  $s^2 = \sum_{t:s_t=k} (y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}})^2$ .
- (8) **Sampling hyperparameters  $\alpha, \kappa$  and  $\eta$ :**
- (i) **Sample  $\alpha + \kappa$ :**
- (a) For  $k = 1, \dots, K$ , sample  $v_k$  from  $\varphi(v_k|\alpha + \kappa, \mathbf{s}) = \text{Bernoulli}(\frac{n_k}{n_k + \alpha + \kappa})$ .
- (b) For  $k = 1, \dots, K$ , sample  $w_k$  from  $\varphi(w_k|\alpha + \kappa, \mathbf{s}) = \text{Beta}(\alpha + \kappa + 1, n_k)$ .
- (c) Sample  $\alpha + \kappa$  from  
 $\varphi(\alpha + \kappa|\mathbf{v}, \mathbf{w}, \mathbf{m}) = \text{Gamma}(e_0 + m_{\cdot\cdot} - \sum_{k=1}^K v_k, f_0 - \sum_{k=1}^K \log w_k)$ .
- (ii) **Sample  $\rho$ :** Draw  $\rho$  from  $\varphi(\rho|\mathbf{m}, \mathbf{r}) = \text{Beta}(g_0 + r_{\cdot}, h_0 + m_{\cdot\cdot} - r_{\cdot})$ .
- (iii) **Calculate  $\alpha$  and  $\kappa$ :** Set  $\alpha = (1 - \rho)(\alpha + \kappa)$  and  $\kappa = \rho(\alpha + \kappa)$ .
- (iv) **Sample  $\eta$ :**
- (a) Sample  $\nu$  from  $\varphi(\nu|\eta, \overline{\mathbf{m}}) = \text{Bernoulli}(\frac{\overline{m}_{\cdot\cdot}}{\overline{m}_{\cdot\cdot} + \eta})$ .
- (b) Sample  $\lambda$  from  $\varphi(\lambda|\eta, \overline{\mathbf{m}}) = \text{Beta}(\eta + 1, \overline{m}_{\cdot\cdot})$ .
- (c) Sample  $\eta$  from  $\varphi(\eta|\lambda, \nu, \overline{\mathbf{m}}) = \text{Gamma}(r_0 + \overline{K} - \nu, s_0 - \log \lambda)$ , where  $\overline{K} = \sum_{k=1}^K 1(\overline{m}_{\cdot k} > 0)$ .
- (9) **Repeat (1) - (8).**