IMPLIED DISTRIBUTIONS IN MULTIPLE CHANGE POINT PROBLEMS

J. A. D. ASTON[1], J. Y. PENG[2] AND D. E. K. MARTIN[3]

[1]CENTRE FOR RESEARCH IN STATISTICAL METHODOLOGY, UNIVERSITY OF WARWICK

[2]INSTITUTE OF INFORMATION SCIENCE, ACADEMIA SINICA

[3]DEPT. OF STATISTICS, NORTH CAROLINA STATE UNIVERSITY

ABSTRACT. A method for efficiently calculating exact marginal, conditional and joint distributions for change points defined by general finite state Hidden Markov Models is proposed. The distributions are not subject to any approximation or sampling error once parameters of the model have been estimated. It is shown that, in contrast to sampling methods, very little computation is needed. The method provides probabilities associated with change points within an interval, as well as at specific points.

Date: June 20, 2011.

Key words and phrases. Finite Markov Chain Imbedding; Hidden Markov Models; Change Point Probability; Run Length Distributions; Generalised Change Points; Waiting Time Distributions.

1. INTRODUCTION

This paper investigates some exact change point distributions when fitting general finite state Hidden Markov models (HMMs), including Markov switching models. Change point problems are important in various applications, including economics (Hamilton 1989; Chib 1998; Sims and Zha 2006) and genetics (Durbin, Eddy, Krogh, and Mitchison 1998; Eddy 2004; Fearnhead and Liu 2007). In many instances change point problems are framed as product partition models (Barry and Hartigan 1992; Barry and Hartigan 1993) or HMMs (Chib 1998; Frühwirth-Schnatter 2006; Fearnhead and Liu 2007), however, to date, the characterisation of the change point distributions implied by these models has been mostly performed using sampling methods (Albert and Chib 1993; Cappé, Moulines, and Rydén 2005), exact posterior sampling (Fearnhead 2006; Fearnhead and Liu 2007), or the distributions are ignored and deterministic algorithms such as the Viterbi algorithm (Viterbi 1967) or posterior (local) decoding (Juang and Rabiner 1991) are used to determine the implied change points. Methods have also been suggested for using a forward-backward like algorithm to get efficient samples based on the Viterbi sequence (Guédon 2007), resulting in posterior distributions for change points (Guédon 2008). In contrast to these previous methods, the distributions in this paper are exact (in the applied probability sense) in that, conditioned on the model and parameters, they completely characterise the probability distribution function without requiring any asymptotic or other approximation or being subject to any sampling error. It will be shown that these distributions can be calculated in many cases using a small number of calculations compared to those needed to yield an approximate distribution through sampling.

The method can be used to determine probabilities of whether a change in regime has occurred at any particular time point. This will be evaluated through a concept called the change point probability (CPP), which is a function of the marginal probabilities of particular time ordered change points occurring at certain times. The marginal probabilities are determined after finding the joint and conditional distributions of multiple change points. Using the marginal distributions allows a probabilistic quantification of the relationship between changes in the behaviour described by the model and real-life events occurring at specific times.

A model might be deemed to capture the influence of an event causing a change in the data when the probability distribution of a change point around the time point of the event is peaked. In contrast, a model for which the probability of a change in regime is more uniform indicates that the regime specified by the model is not particularly affected at that or any other particular point. To illustrate this point, US Gross National Product (GNP) regime switches will be examined in relation to the CPP of recession starts and ends as determined by the National Bureau of Economic Research (NBER). The NBER can be seen as providing external estimates of change points based on overall economic data. Comparing change points determined by the NBER and those determined by maximisation of local posterior state probabilities (Hamilton 1989) leads to the surprising result that regime switches determined from the latter method may not be a useful metric. In contrast, the CPP gives the exact probability of a change occurring at any time point or interval, given the model.

HMMs are widely used in statistics and engineering; see MacDonald and Zucchini (1997) and Cappé, Moulines, and Rydén (2005) for good overviews on the current state of the art both in theory and applications of them. To locate change points (or equivalently to perform data segmentation) HMMs are generally trained on data and then applied to test data (Rabiner 1989; Durbin et al. 1998). The methodology of this paper is appealing in that it allows complete quantification of uncertainty in test data analysis. This methodology is generic in that it depends only on the Markovian nature of regime switches and the ability to generate posterior probabilities, and not upon the structure of any particular HMM model.

The structure of the paper is as follows. Section 2 contains the methods to find change point distributions from data via HMMs and the use of waiting time distributions. Also in that section, the joint and marginal distributions of a set of change points are derived and the concept of CPP is defined. Section 3 contains applications of the methodology to Chib's (1998) change point model and the GNP data given by Hamilton (1989). Section 4 contains a few concluding remarks.

In the supplementary material, a basic smoothing algorithm for Markov switching models will be given, completing an algorithm given by Kim (1994) so that the smoother contains all needed terms.

2. WAITING TIME DISTRIBUTIONS AND CHANGE POINTS

The methods that will be presented here can be applied to general finite state Hidden Markov Models (including Markov switching models) with the form:

$$\begin{aligned} y_t &\sim f(S_{t-r:t}, y_{1:t-1}), \quad t = 1, \ldots, n \\ P[S_{t'}|S_{-r+1:t'-1}] &= P[S_{t'}|S_{t'-1}], \quad t' = -r+2, \ldots . \end{aligned} \tag{2.1}$$

The data y_t from time 1 to time n is distributed conditional on previous data and r previous switching states $S_{t-r}, \ldots, S_{t-1}$ in addition to the current state S_t (as well as model parameters, the values of which are implicitly assumed to be fixed). Here, $y_{t_1:t_2} = y_{t_1}, \ldots, y_{t_2}$ with $S_{t_1:t_2}$ defined analogously. For simplicity, the switching states $\{S_t\}$ are assumed to be a first-order Markov chain with finite state space $\mathcal{S}$, but extension to higher-order Markov structures is straightforward. A given initial distribution π for $S_{-r+1:0}$ is also assumed. With suitable modification, the above model may also include exogenous variables. No assumption on the distribution of the data is made other than that the posterior probabilities of the states (probabilities conditional on the data) must exist.

Definition A run of length k in state s is defined to be the consecutive occurrence of k states that are all equal to s, i.e. $S_{t-k+1} = s, \ldots, S_t = s$ (c.f. Feller (1968), Balakrishnan and Koutras (2002)).

Here k will always represent the minimum length of the regime to be considered or equivalently the length of run to be found. Now for $m \geq 1$, let $W_s(k, m)$ denote the waiting time of the mth run of length at least k in state s, and let $W(k, m)$ denote the waiting time for the mth run of length at least k of any state $s \in \mathcal{S}$. Note that $W(k, m)$ is invariant under any state re-labelling, whereas $W_s(k, m)$ is not.

Consider, for example, the case of growing ($s = 1$) and falling GNP ($s = 0$). $W_0(k, 1)$ is then the first time that a period of falling $(\overbrace{0 \dots 0}^{=k})$ GNP has occurred in $\{S_t\}$ whereas $W(k, 1)$ is the first time either a period of falling GNP or growing $(\overbrace{1 \dots 1}^{=k})$ GNP occurs. By changing the value of k, shorter or longer length periods can be investigated.

A change point at time t is typically defined to be any time at which $S_{t-1} \neq S_t$, the beginning of a run of length at least one. A special case is given in Chib (1998), where the states are required to change in ascending order. However, a more general definition is allowed here, where a change point is defined to have occurred when a change persists at least k time periods, $k \geq 1$. A classic example of when the generalised definition is needed is the common definition of a recession, where two quarters of decline are required ($k = 2$) before a recession is deemed to be in progress. Let $\tau_i^{(k)}$, $i = 1, \dots, m$ be the time of the ith change point under this generalised definition. Then

$$P[\tau_i^{(k)} = t] = P[W(k, i) = t + k - 1]. \tag{2.2}$$

Equation (2.2) follows because the ith run of length at least k occurs at time $t+k-1$ if and only if the switch into that regime has occurred $k - 1$ time points earlier. When $k = 1$, it is assumed in this work that a change point has occurred at $t = 1$, i.e. $P[W(1, 1) = 1] = P[\tau_1^{(1)} = 1] = 1$, and hence the ith change point using the common definition will be equivalent to $(i + 1)$st change point $\tau_{i+1}^{(1)}$ as defined here. Since the Markov chain is assumed to continue even after the observed data, a regime or run of length at least k can appear even after time n and hence $P[\tau_{i+1}^{(k)} > n] \geq P[\tau_i^{(k)} > n] \geq 0$.

Other distributions can be calculated from the waiting time distribution. For example, the distribution of the maximal length of a regime in s, $R_s(t)$, up to time t is given by

$$\begin{aligned} P[R_s(t) = k] &= P[R_s(t) \geq k] - P[R_s(t) \geq k + 1] \\ &= P[W_s(k, 1) \leq t] - P[W_s(k + 1, 1) \leq t]. \end{aligned} \tag{2.3}$$

Analogously, the probability that the maximal length of any regime is k, $P[R(t) = k]$, can be defined in terms of $P[W(k, 1) \leq t]$. In addition, the number of regime changes into state s of length at least k, $N_s(k)$, and the number of change points $N(k)$, can be defined similarly.

2.1. **Methods to Calculate Waiting Time Distributions for the First Change Point.** Finite Markov chain imbedding will be used to compute distributions associated with the regime periods. These methods have been used previously for Markovian problems (Fu and Koutras 1994) and HMMs with discrete observations (Aston and Martin 2007) where it was assumed that there was no dependence between observations except through the underlying Markov chain. The method presented here involves imbedding the $\{S_t\}$ sequence into a new Markov chain $\{Z_t\}$ with a larger state space. Though $\{S_t\}$ forms a homogeneous Markov chain, conditioning on the data induces rth order inhomogeneous dependence, i.e. the posterior transition probabilities $P[S_t|S_{t-r:t-1}, y_{1:n}]$ are transition probabilities of an inhomogeneous r-th order Markov process (Cappé, Moulines, and Rydén (2005) and supplementary material).

The state space of $\{Z_t\}$ (which is denoted by $\mathcal{Z}_s$ or $\mathcal{Z}$ depending on whether runs of a particular state s or a run of any state is of interest) will consist of vector states of the form $((s_{t-r+1}, \ldots, s_t), j)$. The component $(s_{t-r+1}, \ldots, s_t) \in \mathcal{S}^r$, necessary due to the rth-order dependence of states conditional on the data $y_{1:n}$, gives the values of the last r states at time t, $t = 0, \ldots, n$. The component j, $j = 0, 1, \ldots, k$, gives the length of the current run of a particular state s ($j = \max_{1\leq\phi\leq k} : S_t = s, S_{t-1} = s, \ldots, S_{t-\phi+1} = s$ if $S_t = s$, $j = 0$ otherwise), or of the current value of S_t if general runs are of interest. If $k > r$,

$$\mathcal{Z}_s = \bigcup_{j=0}^{r-1} \left\{ \bigcup_{s_{t-r+1:t}:s_l=s, l=t-j+1,\ldots,t;\, s_l\in\mathcal{S}, l=t-r+1,\ldots,t-j} (s_{t-r+1:t}, j) \right\} \cup \left(\bigcup_{j=r}^{k} ((s, \ldots, s), j) \right), \tag{2.4}$$

and if $k \leq r$,

$$\mathcal{Z}_s = \bigcup_{j=0}^{k} \left\{ \bigcup_{s_{t-r+1:t}:s_l=s, l=t-j+1,\ldots,t;\, s_l\in\mathcal{S}, l=t-r+1,\ldots,t-j} (s_{t-r+1:t}, j) \right\}, \tag{2.5}$$

where any strings $s_{a:b}$ with $a > b$ or any s_b with $b < t - r + 1$ are ignored. (Notice that in (2.4) and (2.5), some states are needed only for the initialisation stage when $t < r$).

When $j = k$, a run of length k or longer has occurred. The set A of states with $j = k$ are absorbing, i.e., once entered, the sequence, Z_t, remains in that set of states with probability

one. The state space $\mathcal{Z}$ for calculating $P[W(k,1) \leq t]$ is then

$$\mathcal{Z} = \bigcup_{s \in \mathcal{S}} \mathcal{Z}_s. \tag{2.6}$$

Let z^* represent the size of either $\mathcal{Z}_s$ or $\mathcal{Z}$ as appropriate. As the components of states of $\{Z_t\}$ are functions of states $\{S_t\}$, the $|S|$ non-zero row entries in the $z^* \times z^*$ transition probability matrix M_t for transitions from transient states of $\{Z_t\}$ are completely determined by the posterior transition probabilities $P[S_t|S_{t-r:t-1}, y_{1:n}]$. Specifically,

$$\begin{aligned} &P[Z_t = ((s_{t-r+1}, \ldots, s_t), j) | Z_{t-1} = ((s_{t-r}, \ldots, s_{t-1}), l), y_{1:n}] \\ &= P[S_t = s_t | S_{t-1} = s_{t-1}, \ldots, S_{t-r} = s_{t-r}, y_{1:n}] \end{aligned} \tag{2.7}$$

for appropriate values of j which are determined in the following manner: for transient states of $\mathcal{Z}_s$, $j = l + 1$ when $s_t = s$, and $j = 0$ if $s_t \neq s$. For transient states of $\mathcal{Z}$, $j = l + 1$ when $s_t = s_{t-1}$, and $j = 1$ if $s_t \neq s_{t-1}$. In the supplementary material, a completion of Kim's algorithm is given for the purpose of calculating the posterior transition probabilities of (2.7).

The initial probability distribution for Z_0 is contained in the $1 \times z^*$ row vector ψ_0, which has non-zero probabilities

$$\psi_0((s_{-r+1}, \ldots, s_0), 0) = P[Z_0 = ((s_{-r+1}, \ldots, s_0), 0)] = \pi(s_{-r+1}, \ldots, s_0). \tag{2.8}$$

From the well-known Chapman-Kolmogorov equations for Markov chains (Feller 1968), it follows that the $1 \times z^*$ probability vector ψ_t of Z_t lying in its various states at time $t \geq 1$ is given by

$$\psi_t = \psi_0 \prod_{l=1}^{t} M_l. \tag{2.9}$$

The waiting time distribution $P[W_s(k,1) \leq t]$ can then be calculated as

$$P[W_s(k,1) \leq t] = P[Z_t \in A] = \psi_t U(A), \tag{2.10}$$

with the analogous result holding for $P[W(k,1) \leq t]$, where $U(\Omega)$ is a $z^* \times 1$ column vector with ones in the locations of the set of states Ω and zeros elsewhere. Equation (2.10) holds

since the Markov chain $\{Z_t\}$ is in an absorbing state if and only if a run of length at least k has occurred. Combining (2.9) and (2.10), for $t \geq k$,

$$P[W_s(k,1) = t] = P[W_s(k,1) \leq t] - P[W_s(k,1) \leq t-1] = \psi_0 \left(\prod_{l=1}^{t-1} M_l \right) (M_t - I)U(A), \tag{2.11}$$

where I is a $z^* \times z^*$ identity matrix.

2.2. **Methods to Calculate Waiting Time Distributions for Multiple Change Points.** In this subsection, a method is given to calculate joint probabilities associated with change points through augmenting the state spaces $\mathcal{Z}_s$ and $\mathcal{Z}$. Manipulations of the joint probabilities will lead to an algorithm for computing marginal change point distributions. The algorithm obviates the need to repeat states for each of the $i = 1, \ldots, m$ change point occurrences.

2.2.1. *Setup of Markov Chain for Distributions Associated with Multiple Change Points.* A set of states C, called continuation states, is added to $\mathcal{Z}_s$ and $\mathcal{Z}$, and the respective sizes z^* are incremented by the number of continuation states. The role of the continuation states is that once the ith run of length at least k has occurred, a new Markov chain $\{Z_t^{(i+1)}\}$ is started to determine probabilities associated with the next occurrence of a run of the desired length. The continuation states serve to initialise the new chain $\{Z_t^{(i+1)}\}$, and indicate that run i is still in progress and needs to end before the $(i+1)$st run can begin.

The continuation states $((s_{t-r+1}, \ldots, s_t), -1) \in C$ correspond to absorbing states $((s_{t-r+1}, \ldots, s_t), k) \in A$, with -1 indicating that a run continues and must end for the next run to begin. The (less than full rank) $z^* \times z^*$ matrix Υ defined by

$$\Upsilon\left(z_1, z_2\right) = \begin{cases} 1 & \text{if } z_1 \in A \text{ and } z_2 \in C \text{ is the corresponding continuation state} \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

maps probabilities of being in the states of A into probabilities for being in the corresponding states of C.

The transition probability matrices M_t are revised to account for the continuation states. Continuation states may only be entered from other continuation states. The generic non-zero transition probabilities beginning in a continuation state $((s_{t-r+1}, \ldots, s_t), -1) \in C$, and

conditional on the data are of the form

$$P[Z_t = ((s_{t-r+1}, \ldots, s_t), j) | Z_{t-1} = ((s_{t-r}, \ldots, s_{t-1}), -1), y_{1:n}]$$

$$= P[S_t = s_t | S_{t-r} = s_{t-r}, \ldots, S_{t-1} = s_{t-1}, y_{1:n}], \tag{2.13}$$

where the appropriate values of j for (2.13) are determined by:

(1) If $s_t = s_{t-1}$, $j = -1$ for both $\mathcal{Z}_s$ and $\mathcal{Z}$;

(2) If $s_t \neq s_{t-1}$, then $j = 0$ for $\mathcal{Z}_s$, and $j = 1$ for $\mathcal{Z}$,

with the last $j = 1$ following as runs in any state are of interest. The transition probabilities for the rest of the states in either $\mathcal{Z}_s$ or $\mathcal{Z}$ are unchanged.

2.2.2. *Computation of Joint, Conditional and Marginal Distributions.* The joint distribution of the first m change points can be factorised as

$$P[\tau_m^{(k)} = t_m, \ldots, \tau_1^{(k)} = t_1] = P[\tau_1^{(k)} = t_1] \prod_{i=2}^{m} P[\tau_i^{(k)} = t_i | \tau_{i-1}^{(k)} = t_{i-1}, \ldots, \tau_1^{(k)} = t_1], \tag{2.14}$$

which will equal zero if the distance between any t_i and t_{i-1} is less than k. If $\left(\prod_{l=a}^{b} M_l\right) = I$ for $b < a$, then by (2.11),

$$P[\tau_1^{(k)} = t_1] = P[W(k,1) = t_1 + k - 1] = \psi_0 \left(\prod_{l=1}^{t_1+k-2} M_l \right) (M_{t_1+k-1} - I) U(A). \tag{2.15}$$

To calculate $P[\tau_i^{(k)} = t_i | \tau_{i-1}^{(k)} = t_{i-1}, \ldots, \tau_1^{(k)} = t_1]$, define

$$\xi_{t_1+k-1}^{(1)} = \psi_0 \left(\prod_{l=1}^{t_1+k-2} M_l \right) (M_{t_1+k-1} - I). \tag{2.16}$$

and for $i = 2, \ldots, m$

$$\xi_{t_i+k-1}^{(i)} = \left(\frac{\xi_{t_{i-1}+k-1}^{(i-1)} \Upsilon}{\xi_{t_{i-1}+k-1}^{(i-1)} U(A)} \right) \left(\prod_{l=t_{i-1}+k}^{t_i+k-2} M_l \right) (M_{t_i+k-1} - I). \tag{2.17}$$

By (2.15) and (2.16), $P[\tau_1^{(k)} = t_1] = \xi_{t_1+k-1}^{(1)} U(A)$. The vectors $\left(\frac{\xi_{t_{i-1}+k-1}^{(i-1)} \Upsilon}{\xi_{t_{i-1}+k-1}^{(i-1)} U(A)} \right)$ serve as the initial distribution for the excursion of the Markov chain $\{Z_t^{(i)}\}$ beginning in a continuation state at time $t_{i-1} + k - 1$, analogous to ψ_0 for the first chain at time zero, and

$$P[\tau_i^{(k)} = t_i | \tau_{i-1}^{(k)} = t_{i-1}, \ldots, \tau_1^{(k)} = t_1] = \xi_{t_i+k-1}^{(i)} U(A). \tag{2.18}$$

Thus, for $i = 2, \ldots, m$, the joint probability

$$\begin{aligned} & P[\tau_i^{(k)} = t_i, \ldots, \tau_1^{(k)} = t_1] \\ &= P[W(k,i) = t_i + k - 1, \ldots, W(k,1) = t_1 + k - 1] \\ &= \prod_{q=1}^{i} \xi_{t_q+k-1}^{(q)} U(A) \\ &= \psi_0 \prod_{q=1}^{i-1} \left[\left(\prod_{l=t_{q-1}+k}^{t_q+k-2} M_l \right) (M_{t_q+k-1} - I)\Upsilon \right] \left(\prod_{l=t_{i-1}+k}^{t_i+k-2} M_l \right) (M_{t_i+k-1} - I)U(A), \end{aligned} \tag{2.19}$$

where $t_0 \equiv 1 - k$ for convenience.

Marginal distributions for change point $\tau_i^{(k)}$, or equivalently the marginal waiting time distribution for the ith run occurrence, $i = 2, \ldots, m$, can then be written as

$$\begin{aligned} P[\tau_i^{(k)} = t_i] &= P[W(k,i) = t_i + k - 1] \\ &= \sum_{1 \le t_1 < t_i} \cdots \sum_{t_{i-2} < t_{i-1} < t_i} P[W(k,i) = t_i + k - 1, \ldots, W(k,1) = t_1 + k - 1] \\ &= \sum_{1 \le t_1 < t_i} \cdots \sum_{t_{i-2} < t_{i-1} < t_i} \prod_{j=1}^{i} \xi_{t_j+k-1}^{(j)} U(A) \\ &= \left\{ \sum_{t_{i-2} < t_{i-1} < t_i} \cdots \left\{ \sum_{t_1 < t_2 < t_i} \left\{ \sum_{1 \le t_1 < t_i} \xi_{t_1+k-1}^{(1)} U(A) \right\} \xi_{t_2+k-1}^{(2)} U(A) \right\} \cdots \right\} \xi_{t_i+k-1}^{(i)} U(A) \end{aligned} \tag{2.20}$$

with the marginal distribution $P[\tau_1^{(k)} = t_1]$ given by (2.15).

Equation (2.20) suggests the use of some form of sum-product algorithm (for a definition of the sum-product algorithm see Kschischang, Frey, and Loeliger (2001)) for its calculation. Let $\psi_t^{(i)}$ be row vectors carrying probabilities for the Markov chain $Z_t^{(i)}$, i.e. the joint probability that the $(i-1)$st run has occurred by time t, and that the chain lies in any one of its states at time t (so that $\psi_t^{(1)} = \psi_t$). The marginal distributions are then

$$P[\tau_i^{(k)} = t_i] = P[W(k,i) = t_i + k - 1] = (\psi_{t_i+k-1}^{(i)} - \psi_{t_i+k-2}^{(i)})U(A), \tag{2.21}$$

the probability of being absorbed at time $t_i + k - 1$, and a similar formula holds for $W_s(k,i)$. Two operations need to be carried out to update $\psi_{t-1}^{(i)}$ to $\psi_t^{(i)}$: (1) Due to the Markovian nature of the system, we must multiply by the transition probability matrix M_t, and (2) Absorption

probabilities for the $(i-1)$st run occurrence must be incremented since they give initial probabilities when waiting for the occurrence of the ith run. These operations may be carried out simultaneously for $i = 1, \ldots, m$ by replacing (2.9) with the following matrix computations.

Let Ψ_t, $t = 0, \ldots, n$ be $m \times z^*$ matrices with ith row $\psi_t^{(i)}$. The initial matrix Ψ_0 then has as its first row ψ_0, with the remaining rows being composed of zeroes since the probability is zero that a run has occurred at time $t = 0$. The algorithm for $t = 1, \ldots, n$ is

$$\Psi_t = \Psi_{t-1} M_t, \tag{2.22}$$

$$\psi_t^{(i)} \leftarrow \psi_t^{(i)} + \psi_{t-1}^{(i-1)}(M_t - I)\Upsilon, \quad i = 2, \ldots m, \tag{2.23}$$

where (2.22) is related to computing the matrix product $\left(\prod_{j=t_{i-1}+k}^{t_i+k-2} M_j\right)$ of (2.19) while (2.23) is related to computing $(M_{t_{i-1}+k-1} - I)\Upsilon$, $i = 2, \ldots m$.

Even though the algorithm is non-linear, it can be carried out in linear time w.r.t. t, as the non-linear update step (2.23) is just a simple alteration to entries in the matrix Ψ_t, requiring only linear time computations.

Using the calculations given above, the distribution of the number of regime changes $P[N_s(k) = i]$ into a particular state is given by

$$P[N_s(k) = i] = P[W_s(k,i) \leq n] - P[W_s(k,i+1) \leq n], \quad i = 0, \ldots, \zeta + \lfloor \frac{n-(k+1)\zeta}{k} \rfloor, \tag{2.24}$$

where $\lfloor x \rfloor$ indicates the integer part of x and $\zeta = \lfloor \frac{n}{k+1} \rfloor$. In practice, the value at which $P[W_s(k,i) \leq n]$ becomes negligible will be $i \ll \zeta + \lfloor \frac{n-(k+1)\zeta}{k} \rfloor$. Analogous results hold for probabilities $P[N(k) = i]$ associated with the number of change points in the data, by considering $P[W(k,i) \leq n]$ for $i = 0, \ldots, \lfloor n/k \rfloor$.

By using the setup above it is also possible to determine distributions associated with the end of a regime (the first time point after a run of length at least k has occurred where the process is no longer in the run state). Let $W_s^e(k,i)$ be the time that the ith run in state s ends, with $W^e(k,i)$ being defined in an analogous fashion for the ith run over all states. Regime i ends

when $Z_t^{(i+1)}$ leaves the continuation states. Thus for $t = 1, \ldots, n-1$,

$$P[W_s^e(k,i) = t] = \psi_t^{(i+1)} \left(I - M_{t+1}\right) U(C), \quad i = 1, \ldots, m-1. \tag{2.25}$$

again with an analogous result for $P[W^e(k,i) = t]$.

2.3. **Change Point Probability.** Changes in regime are often deemed qualitatively to coincide with external events such as the start or end of a recession, or a political or historical event such as the Oil Crisis or September 11. Change point probabilities (CPPs) quantify the chance that a switch occurs at a particular time point or within a particular interval. Since only one regime switch can occur at any particular point, a CPP at time t may be computed by summing the probability of the ith change point occurring or ending at that time over i:

$$\text{CPP}_s(t,k) = \sum_i P[W_s(k,i) = t + k - 1] \tag{2.26}$$

and

$$\text{CPP}_s^e(t,k) = \sum_i P[W_s^e(k,i) = t], \tag{2.27}$$

with analogous definitions for probabilities $\text{CPP}(t,k)$ and $\text{CPP}^e(t,k)$ of change points associated with any state of $\mathcal{S}$.

When $k = 1$ and $t < n$, $\text{CPP}^e(t,1) = \text{CPP}(t+1,1)$, since the end of one regime guarantees the start of another. However when $k > 1$, this is not necessarily the case, as it can take more than one time period before a new regime of length at least k appears.

Probabilities $\text{CPP}(t_1 : t_2, k)$ that at least one change point occurs in $t_1, t_1 + 1, \ldots, t_2$ may be computed using the framework of algorithm (2.22-2.23). If $\psi_t^{(i)}(z)$ is the component of $\psi_t^{(i)}$ corresponding to state z, then, since $\sum_{z \in \mathcal{Z} \setminus A} \left[\psi_t^{(1)}(z)\right] + \sum_{z \in A} \left[\psi_t^{(1)}(z)\right] = 1$ and $\sum_{z \in \mathcal{Z}} \left[\psi_t^{(i)}(z)\right] = \sum_{z \in A} \left[\psi_t^{(i-1)}(z)\right]$, $i = 2, \ldots, m$,

$$\sum_{i=1}^{m} \sum_{z \in \mathcal{Z} \setminus A} \psi_t^{(i)}(z) + \sum_{z \in A} \psi_t^{(m)}(z) = 1. \tag{2.28}$$

Let $\tilde{\psi}_t(z) = \sum_i \psi_t^{(i)}(z)$, for $z \in \mathcal{Z} \setminus A$. Then

$$\text{CPP}(t_1 : t_2, k) = \tilde{\psi}_{t_1 - 1} \left(\prod_{l=t_1}^{t_2} M_l \right) U(A) \tag{2.29}$$

in an analogous fashion to (2.11), and similarly for $\text{CPP}_s(t_1 : t_2, k)$. By a similar argument and taking care with the continuation states C, $\text{CPP}^e_s(t_1 : t_2, k)$ and $\text{CPP}^e(t_1 : t_2, k)$ can also be found. By (2.28), $\tilde{\psi}_{t_1-1}$ is not an initial distribution since its elements sum to $1-\sum_{z\in A}\psi_t^{(m)}(z)$. However, (2.29) still gives the exact probability.

2.4. **Computational Considerations.** Given the prevalence of Bayesian techniques in the analysis of Markov switching models (see Frühwirth-Schnatter (2006) for a comprehensive review on these techniques), it is of interest to compare the computational cost of calculating the waiting time and change point distributions through the exact scheme above versus drawing samples from the posterior distribution of the states. Of course, in terms of error for fixed parameters, the two approaches cannot be compared as the exact distribution is not subject to any sampling error.

Either when drawing a conditional sample of the underlying states or when calculating the exact scheme, a pass through a Markov chain is necessary. This is achieved through repeated matrix multiplication across all time points. Simple samplers repeatedly draw conditional samples in this way making a comparison of the computational cost fairly straightforward. If the size of the state space of the sampling technique is ν_{samp}, and the size of the state space for the exact technique is ν_{exact}, a naive calculation yields a comparative computational complexity of $O(n \cdot \nu^2_{\text{samp}})$ versus $O(n \cdot \nu^2_{\text{exact}} \cdot m)$. However, every state in either approach has at most $|\mathcal{S}|$ possible transition destinations, so assuming efficient sparse matrix multiplication techniques are used, this reduces to $O(n \cdot |\mathcal{S}| \cdot \nu_{\text{samp}})$ versus $O(n \cdot |\mathcal{S}| \cdot \nu_{\text{exact}} \cdot m)$. Hence all that needs to be compared is the size of the state spaces associated with the two techniques to characterise the computational complexity difference to find a given number (m) of change points.

For drawing conditional samples (for example through Markov Chain Monte Carlo), in general, a state space of size $\nu_{\text{samp}} = |\mathcal{S}|^r$ is needed as the order of dependence of the posterior Markov chain for the model given in (2.1) is r. For the exact scheme with $k > r$, the state space $\mathcal{Z}_s$ (the number of states given in (2.4) plus one for the continuation state) is of size

$$\nu_{\text{exact}} = \sum_{l=0}^{r} |\mathcal{S}|^l + (k - r + 1) = \frac{1 - |\mathcal{S}|^{r+1}}{1 - |\mathcal{S}|} + (k - r + 1) < |\mathcal{S}|^{r+1} + (k - r + 1)$$

while for $\mathcal{Z}$, the size needed is at most

$$\nu_{\text{exact}} = |\mathcal{S}|^r + \frac{1-|\mathcal{S}|^{r+1}}{1-|\mathcal{S}|} + |\mathcal{S}|(k-r+1) < |\mathcal{S}|^r(|\mathcal{S}|+1) + |\mathcal{S}|(k-r+1).$$

Thus when $k > r$, if $k \ll |\mathcal{S}|^r$, the ratio $\nu_{\text{exact}}.m/\nu_{\text{samp}}$ is at most $|\mathcal{S}|m$ (and often less) and hence at most $|\mathcal{S}|m$ equivalent sample computations are needed to calculate the marginal waiting time distributions. Of course as k increases, the number of states will increase, but this is only at a linear rate proportional to k.

For $k \leq r$, the size of the state space required for $\mathcal{Z}_s$ and for $\mathcal{Z}$ respectively are

$$\nu_{\text{exact}} = (\sum_{i=r-k}^{r} |\mathcal{S}|^i) + |\mathcal{S}|^{r-k},$$

and

$$\nu_{\text{exact}} = |\mathcal{S}|^r + (\sum_{i=r-k+1}^{r} |\mathcal{S}|^i) + |\mathcal{S}|^{r-k+1}.$$

Comparing the computational cost for a standard change point analysis with $k = 1$ and m change points, the number of computations required to calculate the exact marginal distributions is the same as for drawing $3m$ simulation samples, which is small in the usual case when m is small. Note that with the sampling approach, the precise amount of sampling needed can be difficult to quantify given convergence and approximation issues.

The above calculations were all based on using a naive sampler, whereas modern computational techniques such as those in Carpenter, Clifford, and Fearnhead (1999), discussed for change points in Fearnhead (2006), can lead to significant computational savings. It is difficult to clarify the exact relationship between these samplers and the proposed method as the computational cost depends additively on both the length of the time series and the number of samples to be drawn, rather than multiplicatively as in the naive sampler, and thus comparison will depend on the ratio of the number of samples needed to the length of the time series. While computing a large number of samples would be more efficient using such techniques, as will be seen in Figure 3 a very large number of samples might be needed for good estimation of the high-dimensional latent space.

3. EXAMPLES

3.1. **Chib's Binary Change Point Model.** As a pedagogical example, an analysis using this methodology for Chib's (1998) change point model is now given. The model can be viewed in the above framework with restrictions as to the underlying HMM (see Chib (1998) for details). In particular, the HMM is conditioned to only have non-zero transition probabilities to either the same state or the next ascending state (i.e. $P(S_{t+1} = j|S_t = i) = 0$ if $j \notin \{i, i+1\}$) , and is also conditioned to start in state 1 and to end in state $m+1$ which results in exactly m change points.

One particular example, given in Chib (1998), is that of the binary change point model where $y_t \in \{0, 1\}$, and $y_t \sim$ Bernoulli(ζ_t) where the parameter ζ_t changes within the sequence. Following the example in Chib's paper, y_t was simulated from the process

$$\zeta_t = \begin{cases} \theta_1 & t \leq 50 \\ \theta_2 & 50 < t \leq 100 \\ \theta_3 & 100 < t \leq 150 \end{cases}$$

with $\theta_1 = 0.5$, $\theta_2 = 0.75$ and $\theta_3 = 0.25$.

An identical setup was used as in Chib (1998) for the model and parameter estimation (with this information also given in the supplementary material) and it can be seen that very similar state estimates occur (see the supplementary figure). Here, in order to obtain precise distributions from state estimates for comparison with the proposed change point methodology in this paper, the MCMC chain was run for 5×10^4 iterations after an initial burn-in period of 1000, rather than 5000 iterations (the number in the original paper), which is adequate for parameter estimation. Computations for the MCMC were performed using R (R Development Core Team 2011) and MCMCpack (Martin, Quinn, and Park 2011).

In this example, $m = 2$, and thus the state space $\mathcal{S} = \{1, 2, 3\}$. As this model is designed to determine change points without restriction on length, k was chosen to be 1. Given the structured nature of the possible transitions in the model, the extended state space $\mathcal{Z} = \{1, 3, A, C\}$, where A and C represent an absorbing and continuation state respectively. Due to the three

state nature of the model, a change into state 2 represents the first change point (i.e. transitioning into state A) while being in state 2 is equivalent to being in the continuation state C, and thus even though state 2 is in $\mathcal{S}$, it is not needed in $\mathcal{Z}$. The required posterior transition probabilities for this model can easily be found using the results from the original Chib paper. The analysis given here is based on the posterior means of the parameters, and compared to the sampled state sequences (which incorporate parameter uncertainty). As can be seen in Figure 1, the change point distributions for the first and second change points (blue and red respectively) are almost identical using the posterior mean parameters (top plot, not incorporating parameter estimation error) and the sampled state sequence (bottom plot, incorporating parameter estimation error). The only area where small differences occur is in the region 20-30 time steps where there is a very small probability of a change (erroneous given the change is at time 50) in the plots which incorporate parameter estimation error. In addition, a heuristic approximation to the change point distribution incorporating parameter uncertainty can be found by using each set of posterior parameters from the MCMC output and then estimating the change point distribution using the methodology above. The pointwise mean of all these change point distributions (5×10^4, one for each parameter set in the MCMC run) can then be found, and this is given in Figure 1 (middle). This is somewhat heuristic given that the parameters at each iteration depend heavily on the sampled state sequence at that iteration, so this will affect the distribution, but, as again can be seen, incorporating the uncertainty does not change the distribution noticeably in this instance. In addition in Section 1 of the supplementary information, a similar analysis is given for $\theta_1 = 0.4$, $\theta_2 = 0.5$ and $\theta_3 = 0.6$. In this case, it is much more difficult to separate the segments with only 50 observations per segment due to the closeness of θ_i. This yields considerable parameter estimation variability which then manifests itself in the respective change point distributions, showing that differences can occur when parameter estimation error is not accounted for by using parameters based for example on the posterior mean. However, again by using the heuristic approach based on the proposed methodology, the distribution is recovered well. Thus, it would be of considerable interest for further work to

find a rigourous methodology to incorporate this parameter uncertainty to avoid these concerns in general.

[Figure 1 about here.]

3.2. **GNP Analysis.** The Markov switching model, a particular form of HMM that relaxes the assumption of independence between the observed data, is popular in economics. One of the first uses of Markov switching models was to investigate the business cycle. Hamilton (1989) analysed logged and differenced quarterly GNP data for the time period 1951:II to 1984:IV (where Roman numerals indicate which quarter within the year is being considered, I,...,IV). He showed that a two-state mean switching AR(4) is a good model for business cycles. In particular, the change points determined by the Hidden Markov AR(4) model were shown to coincide qualitatively well with all the NBER determinations of recessions. We now examine a quantitative assessment of the model using the proposed methodology.

Hamilton's Markov Switching AR model for GNP is defined as

$$\begin{aligned} y_t &= \alpha_{S_t} + z_t, \\ z_t &= \phi_1 z_{t-1} + \cdots + \phi_4 z_{t-4} + \varepsilon_t, \\ \varepsilon_t &\overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \end{aligned} \tag{3.1}$$

where y_t represents the logged and differenced GNP data, S_t is the underlying state that affects the mean α_{S_t}, z_t is an AR(4) process with parameters $\phi_1, \ldots, \phi_4$ and ε_t is a Gaussian white noise process with zero mean and variance σ^2. In this paper, the underlying states are initialised in the invariant distribution for the chain, and a reduced model is used for $y_1, \ldots, y_4$, so that any ϕ_i associated with $y_{-3}, \ldots, y_0$ is set to zero. This initialisation differs slightly from that in Hamilton (1989), where the first r data points were designated as $y_{-r+1:0}$ and thus the full model was applied to a slightly smaller range of data. The effect of which initialisation is used becomes negligible after the first few time points.

The likelihood function of the model is

$$f(y_t|S_{t-4:t}, y_{1:t-1}) \tag{3.2}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\left[(y_t - \alpha_{S_t}) - \phi_1(y_{t-1} - \alpha_{S_{t-1}}) - \cdots - \phi_4(y_{t-4} - \alpha_{S_{t-4}})\right]^2}{2\sigma^2} \right\}.$$

The state space of $\{S_t\}$ is $\{0, 1\}$, where 0 corresponds to a regime of falling GNP and 1 to a regime of growing GNP. The analysis is conditional on the parameter estimates given in Hamilton (1989), who also conditioned posterior state probabilities on those values.

In Hamilton's analysis, the state at time t is the value that maximises the posterior probability $P[S_t = s_t|y_{1:n}]$, and (implicit) change points into a recession are time points where the state switches from one to zero. This method for determining states, known as local or posterior decoding, (Juang and Rabiner 1991; Durbin et al. 1998) is a position specific maximization algorithm. However, depending on the properties of the transition matrix for S_t, it does not necessarily give a complete posterior state sequence that can occur with positive probability. Determining change points through local decoding neither takes into account the common definition of a recession, which requires at least two quarters of negative growth, nor does it properly account for the exponential number of possible state sequences and their associated probabilities.

In the present work, to calculate distributions of changes into recessions, the imbedding state space $\mathcal{Z}_0$ is constructed as outlined in Section 2. The value of $k = 2$ used here corresponds to the common definition of requiring two quarters of falling GNP for a recession to be confirmed. The posterior transition probabilities $P[S_t = s_t|S_{t-1} = s_{t-1}, \ldots, S_{t-r} = s_{t-r}, y_{1:n}]$ are used in the transition probability matrix of the process Z_t.

[Table 1 about here.]

[Figure 2 about here.]

Figure 2 gives more information about the specific structure of the changes in regime for GNP. The first thing to note is that the third period of falling GNP may occur with reasonable probability in two quite different places. This indicates that the variability in assigning the start and end of the second period is high. The bimodal nature then carries on from this point,

indicating that regimes of falling GNP could begin and end at two different places. A sustained period of growth is indicated from the early 1960's until just before 1970, with little variability in this assessment, which concurs with the NBER economic thinking about the time period. For the remainder of the plots in Figure 2, it can be seen that the distribution peaks are still fairly distinct, meaning that the model is characterising regimes as being either falling or growing without too much overlap. This assessment concurs with economic thinking on the nature of the business cycle, in that high frequency oscillations are not likely, and suggests that the model is still capturing properties of the data even towards the end of the data period. Probabilities for regime periods are only plotted for cases with at least 0.05 total probability. There is only a 0.018 chance of there being ten or more periods of falling GNP in the data, so only the first nine change point distributions are plotted.

[Figure 3 about here.]

Table 1 gives the CPPs according to the model of peaks and troughs of the business cycle occurring at locations determined by the NBER. As can be seen, the model produces a quantitatively good fit for most of the NBER determinations but some NBER peaks are not particularly probable under the model, especially the second and sixth peaks (both less than 1% chance of occurring at the NBER date using the model). These two recessions were of particular interest in the Hamilton (1989) analysis as they were associated with the Suez Crisis and the Iranian Revolution, respectively. The present analysis shows that the NBER recession dates closest to these two events likely do not reflect the immediate effect of the two events. It is also interesting to note that while there are two quarters (6 months) difference between the locations of the NBER peak (1957.III) and the peak in the posterior state probabilities (1957.I) for recession two, and three quarters difference (NBER: 1980.I, Posterior Decoding: 1979.II) for recession six, the probability of recession six starting at the NBER point is higher than the probability of recession two starting at the NBER point. This shows that when trying to determine whether events are explained by the model, exact CPP are better than using distance from the time of a peak in the posterior state probabilities for the event of interest. It is also interesting to note that the NBER troughs for the business cycle are more stable under the model than the peaks. This

can be seen in the graph in Figure 2. In addition to the point estimates, interval probabilities of a change point being at most one quarter different from the NBER dates are also given in Table 1, yielding a confidence interval for the NBER dates in addition to the point probabilities.

Figure 3 shows that the time locations for change points are grouped, although it should be remembered that this plot, unlike those of Figure 2, is not a distribution over time due to the multiple regimes in the data, but rather a graph of the CPP at each time point. The probabilities are moderately peaked, indicating that there are only a few times where change points are likely to occur. It is of interest to note however from both this graph and those in Figure 2 that the fourth change point into a recessionary state as determined by Hamilton's analysis (posterior decoding) occurs between two peaks (at 1969.III). Thus this is actually an unlikely time (0.18 is the corresponding probability; see Figure 3) for the change to have occurred, with it being much more likely to have occurred either just before or just after this date (the NBER peak is just after that at 1969.IV). In addition, in Figure 3, a comparison is given to computing the probabilities of change points using a sampling based scheme (of the form indicated in subsection 2.4). Using an efficient definition of the sample space needed to calculate the exact probabilities, the effective number of comparative computational samples for the exact method is 40 (using $m = 20$). As indicated in Figure 3, even 10^3 samples have some fairly large sampling errors, while 10^5 samples basically capture all the information (although still with a small amount of sampling error (e.g. around 1979)).

In addition to the above analysis, distributions concerning the number of regimes or the length of particular regimes could easily be found (see Section 2 of supplementary material), which could be used to give further interpretation to the properties of the model.

4. Concluding Remarks

In this paper, methods for calculating change point distributions in general finite state HMMs (or Markov switching models) have been presented. A derived link between waiting time distributions and change point analysis has been exploited to compute probabilities associated with change points. The methodology provides a means of improved inference, as change points

determined by maximising the conditional probability at each time point of states given the data can be misleading. As a by-product of deriving the theoretical basis for the approach, smoothing algorithms in the literature have been investigated and a correction to the algorithm given in Kim (1994) has been presented in the supplementary material.

Functions of run distributions have been examined. It would be straightforward to extend the ideas to patterns that are more complex than runs, and to models with multiple regimes. The definition of change points has been generalised to force a sustained change before a change point is counted, however the normal definition of change point (setting $k = 1$ and without the first point being deemed a change point) as a switch from the current regime can be handled by a slight modification of the methodology, namely by adding a continuation state for the zeroth occurrence.

If the distribution of the change points is of fundamental interest, and a suitable prior distribution $\pi(\theta)$ is known for the parameters θ, then a Bayesian approach can be considered:

$$P[\tau_i = t|y_{1:n}] \propto \int_\theta P[\tau_i = t|y_{1:n}, \theta]\pi(\theta)d\theta = \int_\theta P[W(1,i) = t|y_{1:n}, \theta]\pi(\theta)d\theta. \tag{4.1}$$

Numerical integration over the parameter space has been used in change point analysis in the past (Fearnhead and Liu 2007). Equation (4.1) does not include the unknown states explicitly; they only occur through terms that may easily be computed exactly. Thus state sampling is not needed, especially if a suitable importance sampling distribution can be found to compute the integral. Most sampling schemes such as MCMC need the likelihood to include the unknown states, a much higher dimensional problem than when only the change point times are considered.

While the distributions given in this paper can be approximated by sampling the unknown states, there are several disadvantages to this approach. Repeated sampling over the set of unknown states is computationally expensive and introduces sampling error into the computations, whereas in the present work $P[W(1,i) = t|y_{1:n}, \theta], i = 1, \ldots, m$ are calculated exactly, at a computational complexity equivalent to at most drawing $3m$ samples in the standard change point approach. As is apparent in Figure 3, a very large number of samples may be required

to approximate the distribution of this higher dimensional space even when the parameters are known, rendering state sampling approaches unappealing. In addition, the problem of lack of invariance of change point results under state re-labelling is avoided using the method presented in this paper.

The CPP point distribution could alternatively be found, without using the techniques presented here, by finding the posterior distribution of $S_{t-\max(k,r)}, \ldots, S_{t-1}, S_t$ using smoothing techniques similar to the smoother given in the supplementary material, but the computation time would then be exponential in k. Also, the methodology presented in this paper is necessary if change point distributions for each individual occurrence are required, and in that case CPP would require no extra computation other than summing probabilities over time points.

All the analysis and computational steps of this paper have been conditioned on the parameter values in the model. The effect of parameter estimation on the distributions calculated has not been explicitly considered; plug-in MLEs or posterior means have been used, as is often done with other techniques such as local decoding or the Viterbi algorithm. However, in a small stability study using asymptotic MLE distributions and Monte Carlo integration for the GNP data, the CPPs were found to be robust to small fluctuations in the parameter estimates (data not shown). This was also seen to be the case in Figure 1. In addition, in applications where the parameters are determined using training data and the model is then applied to test data, the distributions of the change point locations in the test data are exact, conditional on those parameters. If MCMC or other techniques were available for efficient parameter estimation based on sampling the unknown state sequence (e.g. Chib (1998)), it would indeed be possible to generate posteriors for the parameters and then combine these with the proposed methodology to obtain complete distributions without using the sampled state sequences, which may be slower to converge than the parameter estimates themselves. However, care would need to be taken given the role of the underlying states in generating the posterior parameter estimates.

It should also be noted that many Markov switching models, such as state space switching models (Kim 1994), are analysed through the use of approximations to obtain the smoothed state probabilities. This is done to allow computation in situations where complexity becomes

very large when taking advantage of algorithms such as the Kalman filter and also when r is not a fixed value but is dependent on the data length (Frühwirth-Schnatter 2006), for example in moving average Markov switching models. There have been several suggestions for ways to implement approximations (Kim 1994; Billio and Monfort 1998) to the likelihood function in such cases. These techniques can also be used to approximate the smoothed transition probabilities needed in this paper.

In conclusion, methods for further investigation of change points implied by Hidden Markov models (including Markov switching models) have been presented. The methods are ways of detecting and evaluating change points that are implied by a model. In addition, formulas are given to facilitate evaluation of joint and marginal change point distributions. The widespread use of Markov switching models and change point models in statistics and econometrics, along with hidden state segmentation models used in computer science, will provide many possible applications of this work.

ACKNOWLEDGEMENTS

The authors would like to thank the editor, the coordinating editor and a referee for very helpful comments, including a suggested example, which greatly improved the paper. JA was supported by the EPSRC (UK) and HEFCE through the CRiSM programme grant and by the project grant EP/H016856/1. DM was supported by the National Science Foundation under Grant DMS-0805577.

REFERENCES

Albert, J. H. and S. Chib (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics 11*, 1–15.

Aston, J. A. D. and D. E. K. Martin (2007). Distributions associated with general runs and patterns in hidden Markov models. *Annals of Applied Statistics 1*, 585–611.

Balakrishnan, N. and M. Koutras (2002). *Runs and Scans with Applications*. New York: Wiley.

Barry, D. and J. A. Hartigan (1992). Product partition models for change point problems. *Annals of Statistics 20*, 260–279.

Barry, D. and J. A. Hartigan (1993). A Bayesian analysis for change-point problems. *Journal of the American Statistical Association 88*, 309–319.

Billio, M. and A. Monfort (1998). Switching state space models: Likelihood functions, filtering and smoothing. *Journal of Statistical Planning and Inference 68*, 65–103.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer.

Carpenter, J., P. Clifford, and P. Fearnhead (1999). An improved particle filter for non-linear problems. *IEE proceedings - Radar, Sonar and Navigation 146*, 2–7.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*, 221–241.

Durbin, R., S. Eddy, A. Krogh, and G. Mitchison (1998). *Biological Sequence Analysis*. Cambridge University Press.

Eddy, S. R. (2004). What is a hidden Markov model? *Nature Biotechnology 22*, 1315–1316.

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing 16*, 203–213.

Fearnhead, P. and Z. Liu (2007). Online inference for multiple changepoint problems. *Journal of the Royal Statistical Society, Series B 69*, 589–605.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Volume 1. John Wiley & Sons.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Fu, J. C. and M. V. Koutras (1994). Distribution theory of runs: A Markov chain approach. *Journal of the American Statistical Association 89*, 1050–1058.

Guédon, Y. (2007). Exploring the state sequence space for hidden Markov and semi-Markov chains. *Computational Statistics and Data Analysis 51*, 2379–2409.

Guédon, Y. (2008). Exploring the segmentation space for the assessment of multiple change-point models. Technical Report 6619, INRIA.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Juang, B. H. and L. R. Rabiner (1991). Hidden Markov models for speech recognition. *Technometrics 33*, 251–272.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics 60*, 1–22.

Kschischang, F. R., B. J. Frey, and H.-A. Loeliger (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory 47*, 498–519.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. London, UK: Chapman & Hall.

Martin, A. D., K. M. Quinn, and J. H. Park (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software 42*(9), 1–22.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 257–286.

Sims, C. A. and T. Zha (2006). Were there regime switches in US monetary policy? *American Economic Review 96*, 54–81.

Viterbi, A. (1967). Error bounds for convolution codes and an asymptotically optimal decoding algorithm. *IEEE Transactions on Information Theory 13*, 260–269.

ADDRESS FOR CORRESPONDENCE:

JOHN ASTON

CRISM — TEL: +44-24-7657-4808

DEPARTMENT OF STATISTICS — FAX: +44-24-7652-4532

UNIVERSITY OF WARWICK — EMAIL: J.A.D.ASTON@WARWICK.AC.UK

COVENTRY

UK

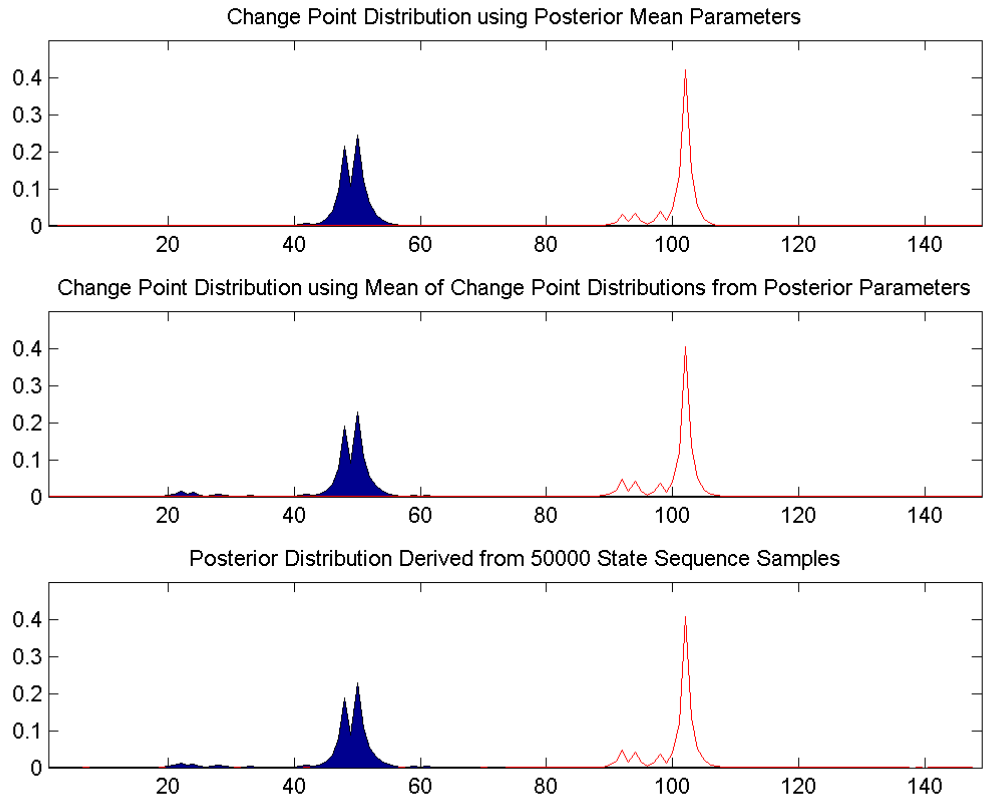


FIGURE 1. Change point distributions calculated using different approaches for Chib's (1998) binary change point model. In each graph the blue area gives the distribution of the time of the first change point and the red line the distribution of the time of the second change point. Top: Distribution of change points calculated using the posterior mean of the parameters. Middle: Distribution of change points calculated by finding the distribution of change points for each parameter set with the MCMC run and then pointwise averaging the distributions. Bottom: Distribution of Change points calculated by determining the distribution from 5×10^4 simulated state sequences.

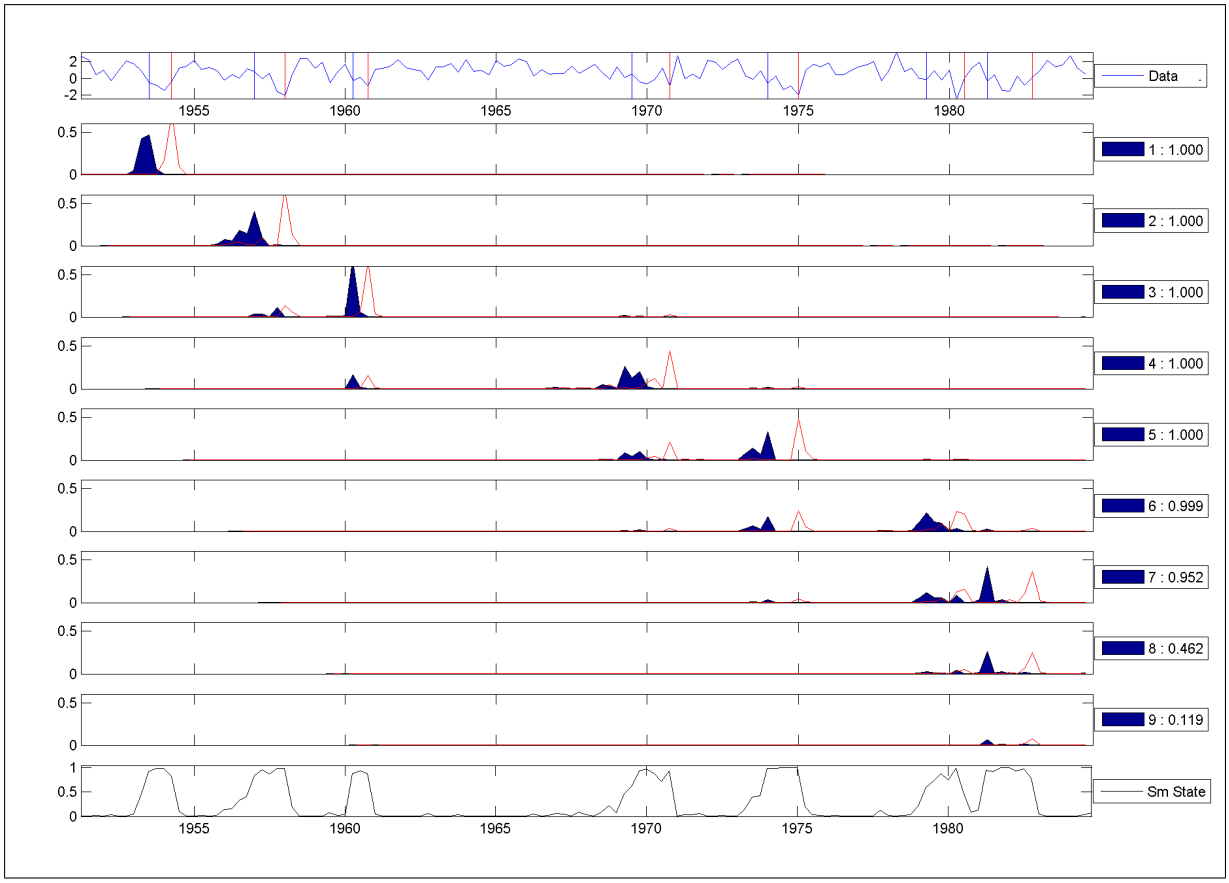
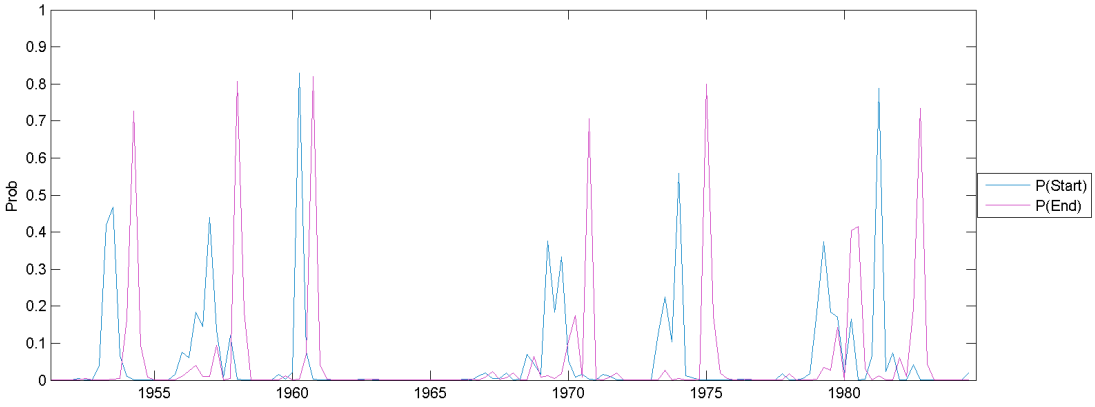
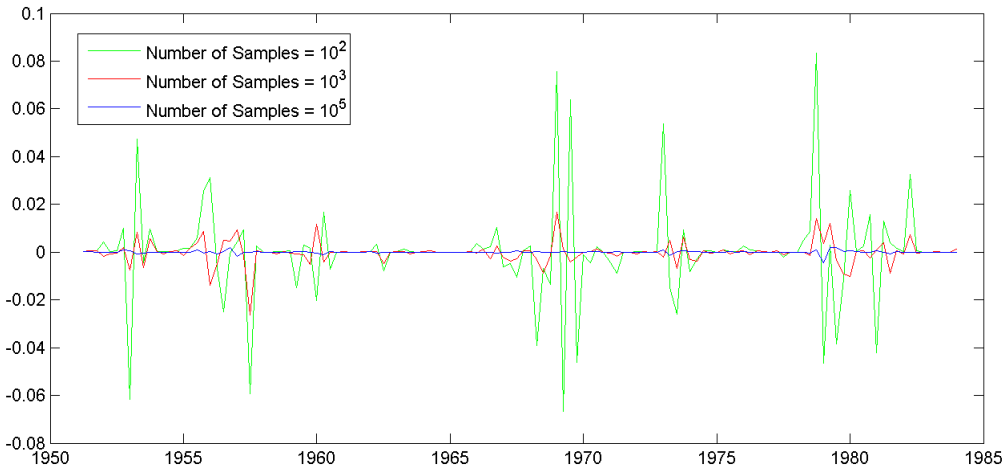


FIGURE 2. Regime Variability of GNP. This is a graphical plot to detect the variability of when different periods of falling GNP states occur in GNP data from 1951:II to 1984:IV. The numbers in the legends to the right of the graphs indicate the index of the period of falling GNP under consideration, followed by the probability of at least that many periods occurring by the end of the data. The very top graph gives a plot of the logged differenced GNP data along with the start (blue line) and finish (red line) of periods where $P[S_t = 0|y_{1:n}] > 0.5$. For the subsequent graphs, the blue area indicates the distribution of the start of a period of falling GNP while the red line indicates the distribution of the end of the falling GNP regime, thus giving a measure of the variability in the length of periods of falling GNP. The ith graph gives the distribution of the ith falling GNP period having occurred at time t, for $i = 1, \ldots, m$. The final plot gives the posterior probability plot $P[S_t = 0|y_{1:n}]$, as used in posterior decoding.



(a) GNP Change Point Probabilities ($k = 2$)



(b) Differences from Exact Probabilities using Sample Based Methods

FIGURE 3. (a) Plot of the CPP of starting (CPP_0) or ending (CPP_0^e) a falling GNP regime within the time period 1951:II to 1984:IV. The graph, while not a distribution, gives information as to the probability of a switch occurring in a particular quarter. (b) This graph indicates the precision of obtaining the same information for CPP_0 through a sampling based estimation scheme for various numbers of samples.

TABLE 1. Dating of the US business cycle peaks and troughs as determined by the NBER, along with their associated probabilities of occurring at or before each time according to the AR(4) mean switching model.

i	Peak (t_1)	$P[W_0(2,i) = t_1+1]$	$\text{CPP}_0(t_1, 2)$	CPP_0 $(t_1-1:t_1+1, 2)$	Trough (t_2)	$P[W_0^e(2,i) = t_2]$	$\text{CPP}_0^e(t_2, 2)$	CPP_0^e $(t_2-1:t_2+1, 2)$
1	1953.III	0.46	0.47	0.93	1954.II	0.72	0.73	0.99
2	1957.III	0.0036	0.0092	0.58	1958.II	0.13	0.18	0.99
3	1960.II	0.66	0.83	0.85	1961.I	0.034	0.042	0.87
4	1969.IV	0.2	0.33	0.89	1970.IV	0.44	0.71	0.72
5	1973.IV	0.065	0.1	0.45	1975.I	0.48	0.8	0.98
6	1980.I	0.0088	0.019	0.37	1980.III	0.2	0.42	0.85
7	1981.III	0.012	0.023	0.88	1982.IV	0.36	0.72	0.96