Exact distributions and Sequential Monte Carlo for Change Point Analysis

Christopher F. H. Nam, John A. D. Aston and Adam M. Johansen,
Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

May 2, 2010

Abstract

Combining recent work on the calculation of exact change point distributions (conditional upon model parameters) via Markov Chain Imbedding and recently developed simulation methodology, this paper presents generally applicable techniques for the estimation of change point distributions in the presence of parameter uncertainty. The proposed approach is both flexible and computationally efficient. Good estimation of the full posterior distribution of the quantities of interest is provided by the proposed methods and this is illustrated via a simulation study and application to GNP data.

Keywords: Finite Markov Chain Imbedding; Hidden Markov Models; Sequential Monte Carlo; Change Points

1 Introduction

Detecting and estimating the location of structural breaks and change points in time series is becoming increasingly important as both a theoretical research problem and a necessary part of applied data analysis. Many parametric and nonparametric approaches to the problem have been proposed using a wide variety of assumptions about the type of breaks of interest and the models for the underlying data outside the breaks (see for example, [14, 3, 4, 11, 15], and references therein). While some of these approaches provide evidence of consistency of

the estimated change points, and others employ sampling-based state space estimation to determine location, few consider explicitly the uncertainty associated with either the number of change points or their location.

In this paper an approach is proposed to characterise uncertainty in the number and location of change points. Conditional on a given set of parameters, exact distributions for the change point number and locations are found, and parameter uncertainty is then incorporated into these distributions through the use of sequential Monte Carlo (SMC). The SMC is only required for the parameter estimates of the segments (without needing to sample the locations as well), and thus the Monte Carlo approximation is on a considerably reduced space. Section 2 summarises the foundations upon which the present work is built; the proposed methodology is then developed in section 3 and applied in section 4. A brief discussion concludes the paper.

2 Background

One particular framework to consider change point problems is through hidden Markov models where the hidden states are associated with the different regimes or segments present in the data. Change points or structural breaks occur whenever there is a change in the underlying state. The methods that will be presented here can be applied to general finite state Hidden Markov Models (including Markov switching models) with the form:

$$\begin{aligned} y_t &\sim f(x_{t-r:t}, y_{1:t-1}, \theta), \\ p(x_t|x_{-r+1:t-1}, \theta) &= p(x_t|x_{t-1}, \theta), \quad t = 1, \ldots, T. \end{aligned} \tag{1}$$

For a given set of parameters θ, the data y_t from time 1 to time T is distributed conditional on previous data and r previous switching states $x_{t-r}, \ldots, x_{t-1}$ in addition to the current state x_t . Here, $y_{t_1:t_2} = y_{t_1}, \ldots, y_{t_2}$ with $x_{t_1:t_2}$ defined analogously. The switching state is assumed to follow a first order Markov chain with finite state space $\mathcal{X}$. A change in the underlying switching state will be associated with a change point, as considered in [14, 3, 10, 11].

2.1 Exact distributions using Finite Markov Chain Imbedding

Conditional on a given set of parameters, it is possible to calculate the exact change point distribution of the model given above. This is done by equating change points with the waiting time distribution of a run. 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. $x_{t-k+1} = s, \ldots, x_t = s$ (c.f. [12], [2]). For $m \geq 1$, let $W_s(k, m|\theta)$ denote the waiting time of the mth run of length at least k in state s, and let $W(k, m|\theta)$ denote the waiting time for the mth run of length at least k of any state $s \in \mathcal{X}$.

When using HMMs, a change point at time t is typically defined to be any time at which $x_{t-1} \neq x_t$, the beginning of a run of length at least one. A special case is given in [3], 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 for at least k time periods, $k \geq 1$. A classic example in which this generalisation is required is the common definition of a recession, in which two quarters of decline are required ($k = 2$) before a recession is deemed to be in progress. Let $\tau_i^{(k)}$, $i = 1, \ldots, m$ be the time of the ith change point under this generalised definition.

By augmenting the state space $\mathcal{X}$ with variables that indicate the progress of a run, in the spirit of finite Markov chain imbedding [13], it is possible to calculate exactly the distribution of $W(k, m|\theta)$ using [1]. The change point distribution can then be found using

$$P[\tau_i^{(k)} = t|\theta] = P[W(k, i) = t + k - 1|\theta]. \tag{2}$$

Equation (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 as in [1].

However, all the above distributions are conditional on θ, which if estimated, as in usual applications, is also subject to estimation error. It is this estimation error that we wish to also incorporate into the change point distribution, through the use of Sequential Monte Carlo samplers.

2.2 Sequential Monte Carlo samplers

In order to deal with parameter uncertainty in a Bayesian framework, we seek to obtain the marginal posterior distribution of the quantities of interest by "integrating out" the parameters themselves. It is not feasible to do this analytically in the models of interest.

Sequential Monte Carlo methods are a class of simulation algorithms which provide samples from a sequence of related distributions by using importance sampling and resampling techniques. Such methods have been well-studied for approximate solution of filtering and related problems in Hidden Markov Models, particularly within the signal processing and econometrics literatures (see [9] and references therein). In fact, such techniques can be employed much more widely and it was shown by [6] that they can be used to provide collections of weighted samples which approximate each of an arbitrary sequence of distributions in turn.

In some situations, it is useful to employ SMC even when a single distribution is of interest. Particularly if this distribution is complex and difficult to sample from or otherwise characterise. In the context of Bayesian inference, if one is interested in sampling from the posterior distribution of θ given a collection of data y, for example, one might consider, for $n = 1, \ldots, P$, a sequence of distributions of the form

$$\pi_n(\theta) \propto p(y|\theta)^{\gamma_n} p(\theta)$$

where $p(\theta)$ corresponds to the prior distribution, $p(y|\theta)$ to the likelihood and $(\gamma_n)_{n=1}^P$ to a non-decreasing sequence, $0 = \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_P = 1$. The intuition is that π_1 is typically easy to sample (or importance sample) from and this sequence of distributions moves smoothly from that to the distribution of interest, $\pi_P(\theta) = p(\theta|y)$.

Markov chain Monte Carlo (MCMC; see [17], for example) methods are the standard for sampling from complex distributions. However, it can be difficult to design MCMC algorithms which mix rapidly enough to provide good samples from the sharply-peaked posterior distributions which arise in many Bayesian problems. SMC algorithms can perform well in this setting as samples do not need to move between modes.

SMC operates by a sequence of mutation and selection operations, in a similar manner to a genetic algorithm. During iteration n, a weighted sample which targets π_{n-1} is mutated

and importance weighted so as to provide a weighted sample which targets π_n. Resampling techniques (loosely, the elimination of samples with small weight and replication of those with larger weights in a systematic fashion which preserves key statistical properties) are also employed to retain stability. Unlike MCMC, SMC employs a population of samples at each iteration and consequently does not require that the mutation step has good global mixing properties.

Considerable freedom exists within the mutation step. For simplicity, we have employed a simple default strategy of using π_n-invariant Markov kernels during iteration n of the algorithm. This approach allows for particularly simple evaluation of the importance weights following the strategy of [6] and works adequately provided that these Markov kernels allow for a reasonable local exploration of the distribution.

3 Methodology

In real problems the model parameters are themselves uncertain. In order to avoid underestimating the uncertainty in the number, location and character of the changepoints this uncertainty must be taken into account. It has not proved possible to obtain analytical expressions for the marginal distribution of the changepoints given the data and prior distributions over the parameters in any degree of generality. Consequently, analyses thus far have consider maximum likelihood estimates of these parameters, together with some sensitivity analysis [1].

When analytical techniques have been exhausted, the next step is typically to invoke numerical integration or simulation-based techniques. Naïve Monte Carlo strategies would proceed via a data augmentation approach. That is, simulate from the joint posterior distribution of the latent state sequence and the parameters of interest, perhaps using MCMC or some other technique to provide these samples. One deficiency of this approach is that it seems wasteful to simulate from such a distribution given that the state sequence is a high-dimensional nuisance parameter and is not needed to calculate the quantities of interest. Furthermore, the latent state sequence is often large as well as highly correlated and it is typically difficult to devise efficient samplers for such situations. That is, not only is effort

expended in estimating a high-dimensional parameter vector in which we are not interested but the inclusion of this parameter vector can make the estimation of the parameters in which we are interested somewhat harder.

The idea that one should use as little Monte Carlo as possible, even when it is necessary to invoke such methods (or that "the only good Monte Carlo is a dead Monte Carlo") dates back at least to [18]. This idea has been widely advocated in the SMC literature in the context of HMM estimation in the guise of Rao-Blackwellised Particle Filters (e.g. [9]) and under other names.

We propose to combine the techniques described in the previous section to provide samples from $p(\theta|y)$ without ever simulating the latent state sequence and then to combine this sample with the exact conditional changepoint distribution to provide a Monte Carlo estimate of the marginal posterior distribution of the quantities of interest.

The key point is that the FMCI approach provides access to the two quantities which are required:

$$p(y|\theta) \text{ and } p(m, \tau_{1:m}|\theta, y)$$

where m is the number of changepoints and $\tau_{1:m}$ is the vector of m changepoint locations.

Given a weighted sample $(W^i, \theta^i)_{i=1}^N$ which is properly weighted to target $p(\theta|y)$ we use the fact that

$$p(m, \tau_{1:m}|y) = \int p(m, \tau_{1:m}|\theta, y)p(\theta|y)d\theta$$

to make the approximation:

$$\widehat{p^N}(m, \tau_{1:m}|y) = \sum_{i=1}^{N} W^i p(m, \tau_{1:m}|\theta^i, y).$$

By expressing $p(m, \tau_{1:m}|y)$ as an expectation, $\mathbb{E}_{p(x|y)}\left[p(m, \tau_{1:m}|X, y)\right]$, under $p(x|y)$ this reduces the estimation of the distribution of interest to the standard Monte Carlo problem of approximating an integral and consequently standard SMC convergence results can be applied (see [5], for an overview).

This is a standard variance reduction technique: any quantity of interest may be cast as an expectation with respect to the joint distribution $p(m, \tau_{1:m}, \theta|y)$ and via the tower

property, for any measurable ϕ:

$$\mathbb{E}[\phi(m, \tau_{1:m})] = E[E[\phi(m, \tau_{1:m})|\theta]]$$

In contrast to more direct Monte Carlo techniques, the proposed method uses the analytic conditional distribution to calculate the inner expectation. It's trivial to show via the law of total variance that using this in place of additional Monte Carlo simulation can only reduce the variance (of an estimator obtained from a sample of fixed size).

In order to specify the SMC sampler it is necessary to specify a number of things, including a mutation kernel which moves the samples obtained at time $n-1$ during iteration n and the importance weighting which depends upon that mutation kernel and the precise extended space construction employed within the sampler. Algorithm 1 provides a general specification which can be employed directly for real problems.

Some explanation of the algorithm's details are probably required. First, resampling is a stochastic procedure by which samples with larger weights are replicated and those with smaller weights are eliminated, the importance weights are set to $1/N$ and the resulting sample remains correctly weighted to target the distribution of interest. There are a number of ways of doing this. The simplest, multinomial resampling, is to sample N times from the weighted empirical distribution but this increases the variance rather more than is necessary. Approaches to resampling were compared in [8].

The purpose of resampling is to stabilise the algorithm by preventing the variance of the importance weights (and hence that of related estimators) from becoming too large. As it increases the Monte Carlo variance it is desirable to avoid resampling too often. The effective sample size (ESS) is a criterion which provides an approximation of the number of independent samples from the target distribution that would be required to provide an estimate of comparable variance which has been used in similar contexts since [16]. The idea being that resampling becomes necessary as the variance of the importance weights increases and the ESS provides a proxy for that unknown variance. Theoretical analysis of algorithms which resample only according to such criteria is provide by [7].

It may seem odd that the algorithm weights and resamples prior to the mutation step. This is possible because the precise construction used in the algorithm leads to importance

Algorithm 1 Outline of the SMC/FMCI changepoint algorithm.

Initialisation:

Sample $\theta_1^i \sim q_1$ for $i = 1 : N$.

For each i, calculate

$$W_1^i = \frac{w_1(\theta_1^i)}{\sum_{j=1}^N w_1(\theta_1^i)} \text{ with } w_1(\theta_1) = fracp(\theta_1)q(\theta_1).$$

If ESS(W_1) $<$ Threshold then resample.

Iteration:

For $n = 2$ to P

For each i, calculate weight

$$W_n^i \propto W_{n-1}^i \times \frac{\pi_n(\theta_{n-1}^i)}{\pi_{n-1}(\theta_{n-1}^i)} \text{ such that } \sum_{i=1}^N W_n^i = 1.$$

If ESS(W_n) $<$ Threshold then resample.

Sample each $\theta_n^i \sim K_n(\theta_{n-1}^i, \cdot)$ where K_n is a π_n-invariant Markov kernel.

End for

Calculation and Output:

Calculate

$$p(m, \tau_{1:m}) = \sum_{i=1}^N W_n^i \, p(m, \tau_{1:m} | y, \theta_n^i).$$

weights which are independent of the new location. Resampling before rather than after mutation is consequently possible and increases the diversity of the resulting sample. This has been discussed previously by, for example, [6].

4 Results

This section presents results obtained by applying the presented methodology on data assumed to be generated from a Markov Mixture model and an Markov switching AR model. We consider a variety of simulated data in the Markov Mixture model case where the change points are fixed and the parameters used to generate the data are known. For the Markov switching AR model, we apply our methodology to the GNP data supplied in [14] in determining the starts and ends of recessions.

4.1 Markov Mixture model

For a Markov Mixture distribution, the distribution of the output is defined as follows

$$y_t|(X_t = i) \sim f_{\theta_i}(\cdot)$$

where θ_i are the parameters associated with the distribution f, $i \in \mathcal{X}$, and X_t is first order Markovian.

4.1.1 Simulated Results

The following datasets are generated from a two state Gaussian Markov Mixture distribution, that is $Y_t \sim \mathcal{N}(\mu_1, \sigma_1^2)$ or $\mathcal{N}(\mu_2, \sigma_2^2)$. In order to determine whether the results are plausible, we fix the underlying Markov Chain and in turn, we are able to determine where the change points are said to have occurred for a particular regime.

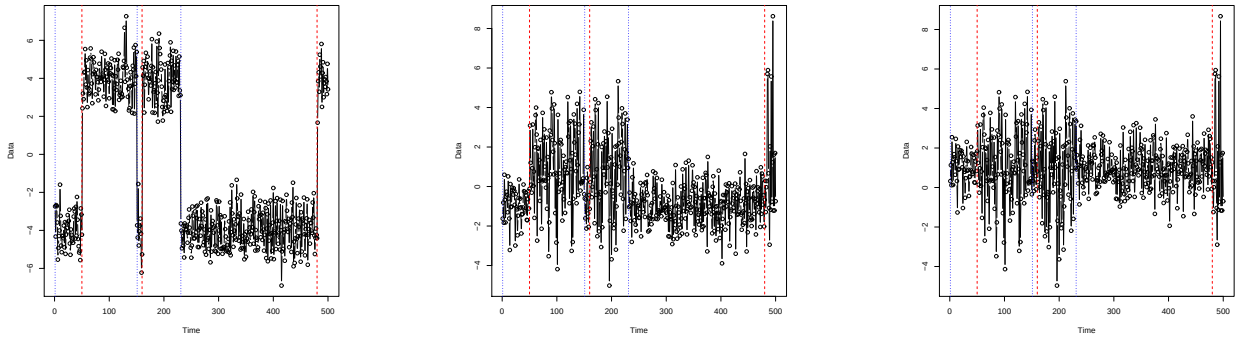
The model parameters of uncertainty are the two means, variances and transition probabilities of the underlying MC. We used 1000 particles and 100 distributions in approximating the posterior of the model parameters. We used as prior distributions the Normal, Gamma and Dirichlet for the means, precisions and transition probabilities respectively. The algorithm was initialised by sampling from these prior distributions during the first iteration. A

Random Walk Metropolis Hastings (RW-MH) strategy was used to propose new parameter values with the precisions and transition probabilities being treated on log-Normal and logit scales, respectively due to the conditions that they must satisfy.

Our fixed underlying MC has the statespace $\{1, 2\}$ and we are particularly interested in the change point switches into the regime consisting of a run of length 5 in state 1. That is $s = 1$, $k = 5$ in this case. Let us refer to this regime of interest as "stable" for simplicity. The data set contains only 3 occurrences of the stable regime. As a comparison, we also compute the exact change point distribution under the posterior mean of the particles, that is the weighted average of the particles.

Figure 1 displays the results of the algorithm to three datasets where it progressively becomes harder to identify the location of the change points by inspection of the data. Our results concur with this as the probability that there were at at least 3 occurrences of stable regimes diminishes (as seen in (a),(b) & (c)) and the probability of the start and ends of the regime become less sharply peaked ((d),(e) & (f))). We observe also, particularly in (f) just before time 100, some other regime starts are signalled, with fairly small non-zero probabilities. This, in turn, results in more than presence of more than three occurences being assigned a small positive probability.

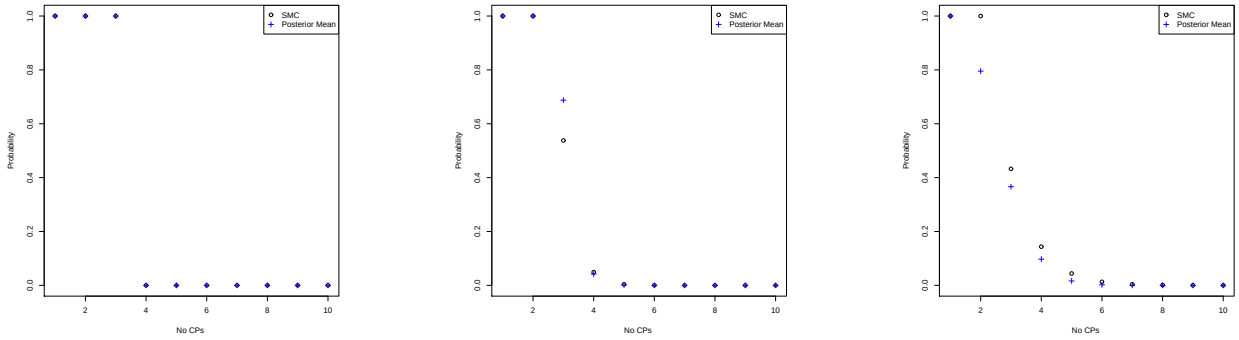
Comparing our approach to an exact distribution approach using a plug-in estimate of the parameters corresponding to the estimated posterior mean, we observe particularly in the 2nd and 3rd column, that the full SMC framework approach allows us to incorporate the general variability amongst all the calculated exact change point distributions. In the 2nd column case, this can be seen with a lower probability for 3 occurrences of the stable regime in (b) for the SMC case. In (e), the probability of the 2nd regime starting around time 150 is slightly less pronounced in the SMC case. In the 3rd column, where the exact change point distribution via posterior means does not seem to detect any of the change points whilst the SMC approach does, (f), this again captures the variability amongst the distributions reflecting that there are parameter values with significant posterior mass which are consistent with these changes. This perhaps highlights the need to calculate a general



Simulated Datasets where the location of change points becomes progressively less apparent

Blue and red dotted lines denote the starts & ends of the regime of interest.

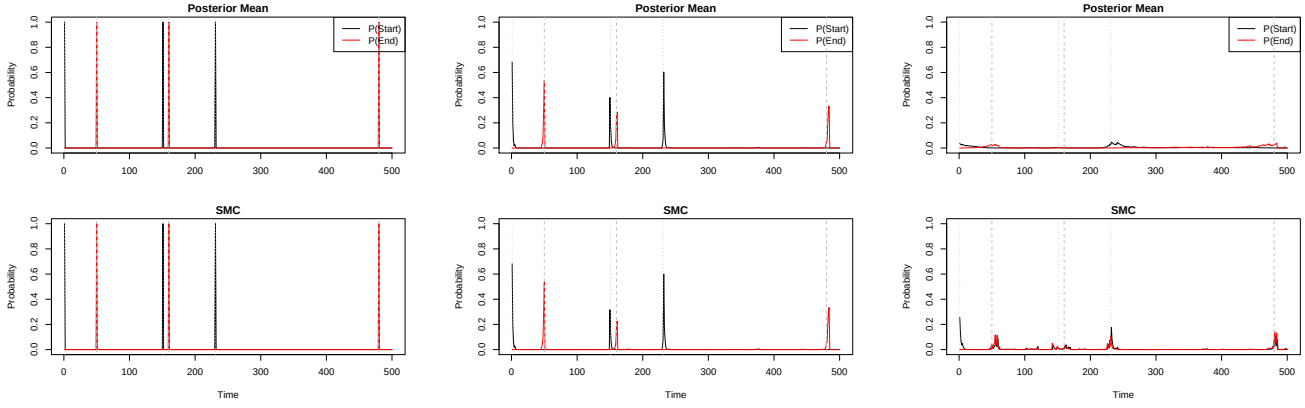
Data: $\mathcal{N}(-4,1) \quad \mathcal{N}(4,1)$ **Data:** $\mathcal{N}(-1,1) \quad \mathcal{N}(1,2^2)$ **Data:** $\mathcal{N}(0.95,1) \quad \mathcal{N}(1.05,2^2)$



(a) Probability of at least w occurrences of stable regime

(b) Probability of at least w occurrences of stable regime

(c) Probability of at least w occurrences of stable regime



(d) Plot of the change point probability of stable starting (black) and ending (red) at each time

(e) Plot of the change point probability of stable starting (black) and ending (red) at each time

(f) Plot of the change point probability of stable starting (black) and ending (red) at each time

Figure 1: The complete methodology applied to a variety of simulated data from a 2 component Markov Mixture distribution. As the separation between the regimes (and their model parameters) becomes less, there is greater uncertainty in the locations of the change points.

change point distribution which incorporates parameter uncertainty as opposed to using a single exact distribution for one parameter setting.

4.2 Markov Switching AR model

For a Markov Switching AR model, y_t is dependent not only upon x_t but also on previous values of $y_{1:t-1}$. When there is a direct dependence on the previous r output values, we denote this as MS-AR(r). A particular case of an MS-AR(r) model is Hamilton's MS-AR(r) model. Hamilton's MS-AR(r) model for observed data y_t is defined as

$$\begin{aligned} y_t - \mu_{x_t} &= z_t \\ z_t &= \phi_1 z_{t-1} + \ldots + \phi_r z_{t-r} + +\epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \end{aligned}$$

where the mean μ switches according to the hidden state x_t. The data without the mean follows a standard autoregressive model of order r, with parameters $\phi_1, \ldots, \phi_r$ and error term ϵ_t, which is assumed to follow a Normal distribution with mean 0 and variance σ^2.

4.2.1 GNP Data

Hamilton's GNP data [14] uses the differenced quarterly logarithmic GNP between the time periods 1951:II to 1984:IV. The data is found to be adequately modelled by the aforementioned Hamilton's MS-AR(4) model where y_t represents the logged and differenced GNP data. The data is of particular interest in determining the start and ends of business cycles, namely when a recession begins and ends.

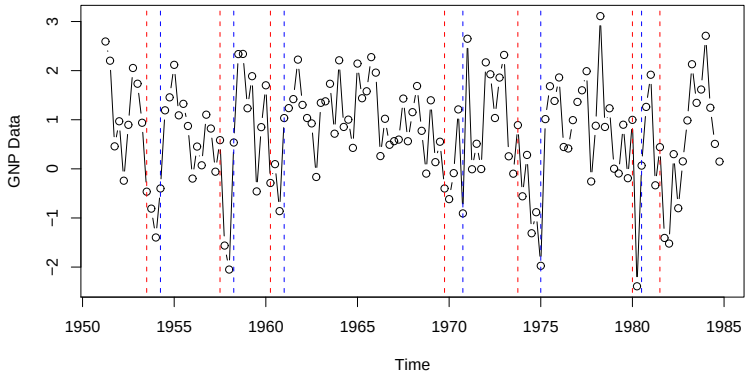
The state space of $\{X_t\}$ consists of two states; 0, corresponding to a regime of falling GNP and 1, corresponding to a regime of growing GNP. A widely held definition of a recession is that it is said to be in progress when there are two consecutive periods of falling GNP; we thus classify a recession to have started when it preceded by a run length of 2 in state 0, that is $s = 0$, $k = 2$. Since the determination of the starts and ends of recession has only been performed qualitatively by NBER (denoted by the red and blue dotted lines respectively in graphs), this makes this an ideal dataset with which to compare our results, and ultimately quantify the uncertainty of these determinations.

The SMC component of our methodology consists in estimating the means associated with each state, the variance and transition probabilities. We use the same priors and proposal setup as outlined in subsection 4.1.1 for the means, variances and transition probabilities. As AR coefficients can be highly influential on the other parameter estimates and thus our final general change point distribution, our primary goal is to examine the effect of AR parameter estimation error on change point distributions.

We firstly constrain the AR coefficients to the maximum likelihood estimates (MLE) as in [14], throughout the entirety of the SMC algorithm, and estimate all other model parameters. This allows an examination of the effect of parameter uncertainty in all parameters except AR parameters. The second approach samples the AR coefficients from a tight prior centered around Hamilton's MLE estimates initially, and proposing moves on these coefficients with a small variance. To ensure stationarity holds, we consider working with the (complex) roots of the AR parameters, and in turn the associated moduli and arguments. A uniform tight prior is used, and RW-MH proposals are performed on the logit scale for the moduli. This thus allows us to account for some slight variation in the AR coefficients, and examine how this affects the resulting general change point distribution.

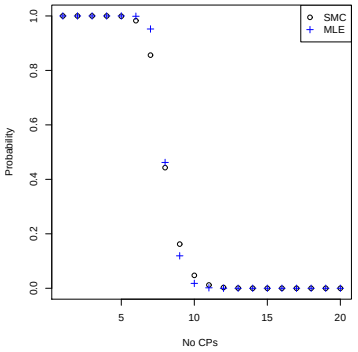
In approximating the posterior distribution of the model parameters under SMC, 1000 particles and 100 time steps were used. We compare our results, the general change point distributions via our methodology under the two discussed cases, with those using Hamilton's MLE estimates where an exact distribution is obtained as no parameter uncertainty is considered.

Both sets of approaches firstly largely concur with the qualitative determinations of recessions starts and ends with the respective change point probabilities for the start and end of regimes being centred and peaked around these times (Figure 2(d) & (e)). In addition, the suggested number of occurrences of recessions (7 according to NBER), occurs with reasonably high probabilities (Figure 2(b) & (c)). In comparison with using the MLE estimates solely, we note that the general output is similar with some features being more pronounced than others in the change point probability graphs (d) & (e), which accounts for the additional



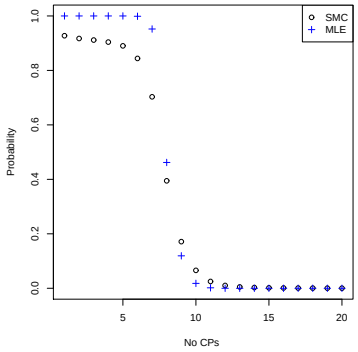
(a) Hamilton's GNP quarterly data from 1951:II to 1984:IV. Red and blue lines denote start and ends of recession respectively according to NBER.

Fixed AR parameters

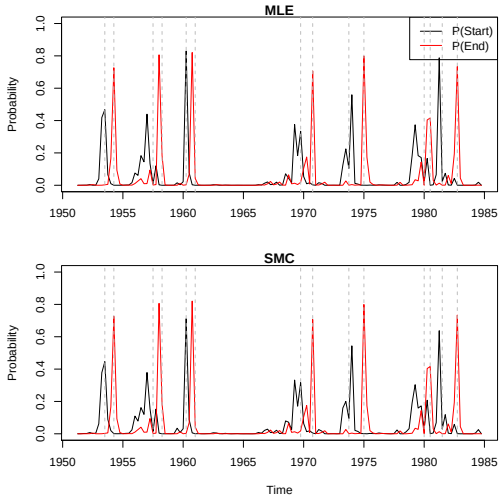


(b) Probability of at least w recessions

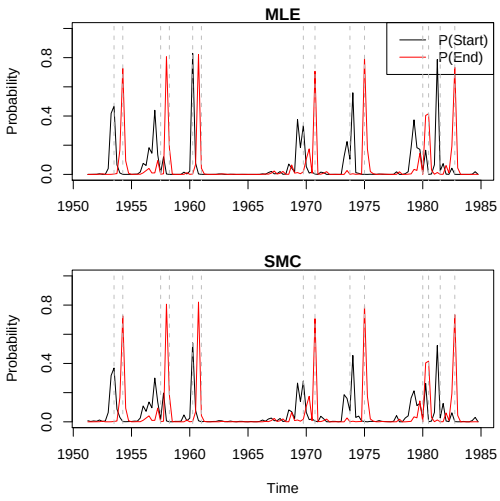
Varying AR priors



(c) Probability of at least w recessions



(d) Plot of the CP probability of a recession starting (black) and ending (red) at each quarter



(e) Plot of the CP probability of a recession starting (black) and ending (red) at each quarter

Figure 2: Methodology applied to Hamilton's GNP data in two cases: fixing the AR parameters (left) and allowing some slight variation in the AR coefficients (right). Both cases concur with NBER qualitative starts and ends of recessions and are similar to results via an MLE approach. Differences between the MLE approach and each other, reflect the uncertainty associated with the parameters and the AR coefficients.

uncertainty from the parameters. The difference between the exact and general distribution is more apparent in the probability of the number of recessions, with the general approach being generally flatter, accounting for the associated uncertainty of the model parameters.

Introducing the slightest variation on the AR coefficients as in our two considered approaches, thus appears to be highly influential on the resultant general change point distribution. As observed in (b) & (c), the uncertainty in the number of recessions drops noticeably when uncertainty in the AR coefficients is introduced, over just having uncertainty in other parameters.

5 Conclusion

The main contribution of this paper is a method for Bayesian estimation of the distribution of change points in the presence of parameter uncertainty. A generic algorithm has been developed, although considerably greater flexibility is available. Proof-of-concept simulation results indicate that the method is effective, while preliminary real data analysis demonstrates that parameter uncertainty cannot be safely ignored. Future work will further develop the simulation framework and discuss the development of simulation algorithms for specific problems in detail.

References

[1] J. A. D. Aston, J. Y. Peng and D. E. K. Martin. Implied Distributions in Multiple Change Point Problems *CRiSM Research Report 08-26*, 2008.

[2] N. Balakrishnan and M. Koutras. *Runs and Scans with Applications*. Wiley, New York, 2002.

[3] S. Chib. Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86:221–241, 1998.

[4] R. A. Davis, T. Lee, and G. Rodriguez-Yam. Structural Break Estimation for Nonstationary Time Series Models. *Journal of the American Statistical Association*, 101, 229-239, 2006.

[5] P. Del Moral *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Series: Probability and Applications, Springer-Verlag, New York, 2004.

[6] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B*, 63(3):411–436, 2006.

[7] P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling procedures for sequential Monte Carlo methods. Technical report. INRIA, 2008.

[8] R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filters. in *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, 2005.

[9] A. Doucet, and A. M. Johansen. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later in *Handbook of Nonlinear Filtering* eds. D. Crisan and B. Rozovsky. Oxford University Press, 2010, to appear.

[10] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis*. Cambridge University Press, 1998.

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

[12] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, 1968.

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

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

[15] C. Kirch *Resampling Methods for the Change Analysis of Dependent Data*. PhD thesis, University of Cologne, Cologne, 2006. http://kups.ub.uni-koeln.de/volltexte/2006/1795/.

[16] A. Kong, J. S. Liu, and W. H. Wong, Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, **89**, 1032–1044, 1994.

[17] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Verlag, New York, second edition, 2004.

[18] H. F. Trotter, and J. W. Tukey, Conditional Monte Carlo for normal samples, in *Symposium on Monte Carlo Methods* (ed. H. A. Meyer) Wiley, New York, 64–79, 1956.