# Bayesian single changepoint estimation in a parameter-driven model

CHIGOZIE E. UTAZI

*Southampton Statistical Sciences Research Institute and WorldPop, Department of Geography and Environment, University of Southampton*

**Abstract**

In this paper, we consider the problem of estimating a single changepoint in a parameter-driven model. The model - an extension of the Poisson regression model - accounts for serial correlation through a latent process incorporated in its mean function. Emphasis is placed on the changepoint characterization with changes in the parameters of the model. The model is fully implemented within the Bayesian framework. We develop a RJMCMC algorithm for parameter estimation and model determination. The algorithm embeds well-devised Metropolis-Hastings procedures for estimating the missing values of the latent process through data augmentation and the changepoint. The methodology is illustrated using data on monthly counts of claimants collecting wage loss benefit for injuries in the workplace and an analysis of presidential uses of force in the US.

## 1   Introduction

Changepoint detection and estimation is a subject of prime importance in the analysis of temporal processes. Changepoints happen when there is an extraordinary event or external influence on the process(es) of interest, causing abrupt or possibly slow changes in the properties of the process(es). Both integer- and

real-valued temporal processes often exhibit changes in behaviour owing to their dynamic structure. Changepoint analysis for temporal processes has a huge statistical literature; see Jandhyala et al. (2013) for an excellent review. Several approaches ranging from the use of CUSUM tests (Kang & Lee, 2014) to maximum likelihood methods (Fu & Curnow, 1990) and Bayesian methods (Carlin et al., 1992; Green, 1995; Barbieri & Conigliani, 1998; Fearnhead, 2006), have been developed for detecting and estimating single and multiple changepoints in both sequential and retrospective formulations.

In the time series regression context, whilst changepoint analysis in linear models with Gaussian errors has received a great deal of attention, incorporation of changepoints into more general parameter-driven time series models is less well-developed and analyzed. In particular, we are not aware of any work that addresses the problem of changepoint detection and estimation in the parameter-driven model proposed by Zeger (1988) and used in Campbell (1994), Chan & Ledolter (1995) and Frühwirth-Schnatter & Wagner (2006). The work of Frühwirth-Schnatter & Wagner (2006) is particularly relevant to this work as their analysis is based upon a data augmentation MCMC which we extend upon to develop a reversible jump MCMC algorithm to detect the presence of a changepoint in the data. The parameter-driven model accounts for autocorrelation and overdispersion through a latent process incorporated in its mean function. This is in contrast to an observation-driven model which explains serial correlation by including lagged values of the variable in its mean function.

The parameter-driven model of Zeger (1988) is a Poisson regression model and this will be the model we focus on in this paper. The approach is readily generalizable to other discrete regression models. Specifically, the Poisson regression model can be written as:

$$Y_t | X_t, \mathbf{z}_t \sim \text{Po}(\exp(\mathbf{z}_t^T \boldsymbol{\beta} + X_t)), \tag{1}$$

$$X_t = \alpha X_{t-1} + \epsilon_t, \tag{2}$$

where $Y_t$ denotes the observed counts, $\mathbf{z}_t$, a covariate vector, $\boldsymbol{\beta}$, the regression parameters and $\{X_t\}$ is an $AR(1)$ process with independent and identically dis-

tributed $\epsilon_t$. Note that when $X_t = 0$, (1) reduces to a standard Poisson regression model. In Zeger (1988) it is assumed that $E[\exp(X_t)] = 1$ to ensure that the model parameters are identifiable and to aid interpretation of the model. That is, this assumption ensures that $E(Y_t) = E(E(Y_t|X_t)) = \exp(\mathbf{z}_t^T \boldsymbol{\beta}) E(\exp(X_t)) = \exp(\mathbf{z}_t^T \boldsymbol{\beta})$. Thus, as is required in the regression context, the unconditional mean of the counts depends only on the linear combination of covariates and not on the moments of the latent process. However, with a stationary $\{X_t\}$, this restriction may be relaxed since the regression coefficients apart from the intercept remain unchanged with varying assumptions about $E(\exp(X_t))$. Also, simple algebraic manipulations can be carried out to make estimates of the intercept consistent across such assumptions (see, for example, Chan & Ledolter, 1995).

The unobserved nature of the latent process $\{X_t\}$ leads to a highly complicated likelihood of the model. In Zeger (1988) an estimating equation approach was used to fit the model. The estimating equation method (Cameron & Trivedi, 2013, p. 279) solves a set of simultaneous equations analogous to the score equations in maximum likelihood theory, and which depend on the first- and second-order moments of the model. However, the form of the parameter-driven model given by (1) means that data-augmentation of $\{X_t\}$ leads to a tractable likelihood. Consequently, various authors have utilised data augmentation for parameter estimation. Chan & Ledolter (1995) developed a Monte Carlo EM algorithm for estimating parameter-driven models. The algorithm is an extension of the EM algorithm as it facilitates the computation of the expectation step using MCMC methods. Oh & Lim (2001) considered a fully Bayesian analysis of the model in which the computational difficulty introduced by the latent process was overcome through a reparameterization of the model in terms of additional Gaussian latent variables. In Frühwirth-Schnatter & Wagner (2006) a general approach based upon Gibbsian transition kernels is introduced which is applicable to a wide range of state space models.

In this paper, we address the problem of detecting and estimating a changepoint in a parameter-driven model. We consider a retrospective formulation - assuming a finite set of data has been collected. We focus on the changepoint parameterization in which a single change is assumed to have occurred in all the

parameters of the model although other parameterizations are possible as mentioned in Section 2. We also use data augmentation (see Section 2.1) to facilitate inference in a Bayesian framework. The key novelty is the introduction of an effective reversible jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995) to jointly estimate the parameters and whether or not a changepoint should be included in the model. While Green (1995) fitted continuous-time changepoint models, we work in discrete time in a modelling framework that is complicated by the presence of missing data.

The rest of the paper is organized as follows. The proposed changepoint model is presented in Section 2. Its Bayesian estimation procedure and the accompanying RJMCMC algorithm are discussed in Sections 2.1 and 2.2, respectively. A simulation study investigating the performance of the model and our algorithm is presented in Section 3. In Section 4, we illustrate the methodology using two data sets: monthly counts of claimants collecting wage loss benefit for injuries in the workplace (see Freeland & McCabe, 2004) and annual counts of instances in which US presidents used force abroad between 1930 and 1995. Finally in Section 5 we make our concluding remarks, in particular, commenting on the extension to multiple changepoint detection.

## 2 Model specification

In this Section we specify the model and detail an MCMC algorithm to efficiently sample from the posterior distribution of the parameters. We then extend the MCMC algorithm to include a reversible jump step to move between the models with and without a changepoint.

Let $\{Y_t : t = 1, \ldots, n\}$ denote the observed counts, $\{X_t : t = 1, \ldots, n\}$, a stationary unobserved latent process, and $\mathbf{z}_t$, a $p$-dimensional covariate vector. For ease of exposition we focus on a latent Gaussian AR(1) process but the extension to other latent processes is possible. Let $m \in \{1, \ldots, n-1\}$ denote the location of a changepoint in the data. Note that taking $m = n$ in the sequel produces a model without a changepoint. For $t = 1, 2 \ldots, m$, let $c_t = 1$ and for $t = m + 1, m + 2, \ldots, n$, let $c_t = 2$. Thus $c_t = 1$ ($c_t = 2$) if observation $t$

4

is before (after) the changepoint. Conditional on the underlying latent process, the deterministic covariate process and the unknown changepoint, the observed counts are assumed to be independently Poisson distributed with mean $\mu_t = \exp(\mathbf{z}_t^T \boldsymbol{\beta}_{c_t} + X_t)$, denoted by

$$
\begin{aligned}
Y_t | X_t, \mathbf{z}_t &\sim \text{Po}(\exp(\mathbf{z}_t^T \boldsymbol{\beta}_{c_t} + X_t)), \\
X_t &= \alpha_{c_t} X_{t-1} + \epsilon_{c_t},
\end{aligned}
\tag{3}
$$

where for $i = 1, 2$, $\epsilon_i \sim N(0, 1/\tau_i)$ and $\boldsymbol{\beta}_i = (\beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,p-1})$ are regression parameters. We assume that $|\alpha_i| < 1$ to ensure stationarity in the AR(1) latent process.

Observe that the model partitions the observed counts and the latent process into two homogenous blocks - one occurring before and the other after some unknown time point, $m$. Given the parameters the properties of the count process such as $E[Y_t]$, $Var(Y_t)$ and $Cov(Y_t, Y_{t+h})$ can easily be obtained, see Zeger (1988) and Davis et al. (2000). Thus allowing study of the behaviour of the count process both before and after the changepoint. The assumption throughout this paper is that all the parameters in the model change at the changepoint. It is straightforward to adapt the following approach to a case where either only the latent process or the regression parameters change.

## 2.1   Prior specification and posterior distribution

For $i = 1, 2$, let $\boldsymbol{\theta}_i = (\boldsymbol{\beta}_i, \alpha_i, \tau_i)$ and let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, m)$. Our interest is in the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$, where $\mathbf{y} = (y_1, \ldots, y_n)$ denotes $n$ observations from the count process $\{Y_t\}$. However, the likelihood $\pi(\mathbf{y}|\boldsymbol{\theta})$ is not tractable due to the need to integrate out the unobserved latent process, $\{X_t\}$. In order to make progress it is necessary to use data augmentation, (see Tanner & Wong, 1987), $\mathbf{x} = (x_1, \ldots, x_n)$ of the latent process within our MCMC framework. The

augmented data likelihood function of the model is given by

$$\pi(\mathbf{x}, \mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) \propto \left\{ \prod_{t=1}^{n} \left[ \exp\left( \mathbf{z}_t^T \boldsymbol{\beta}_{c_t} y_t + x_t y_t - \exp(\mathbf{z}_t^T \boldsymbol{\beta}_{c_t} + x_t) \right) \right] \right.$$

$$\left. \times \prod_{t=2}^{n} \left[ \tau_{c_t}^{\frac{1}{2}} \exp\left( -\frac{\tau_{c_t}}{2}(x_t - \alpha_{c_t} x_{t-1})^2 \right) \right] \right\}$$

$$\times \sqrt{\tau_1(1 - \alpha_1^2)} \exp\left( -\frac{\tau_1(1 - \alpha_1^2)}{2} x_1^2 \right). \tag{4}$$

Note that the final term in (4) follows from $X_1 \sim N(0, 1/\{\tau_1(1 - \alpha_1)^2\})$.

We assume that the parameters are a priori independent and for $i = 1, 2$, choose $N(u_\beta, 1/v_\beta)$ priors for $\beta_{i,j}$ $(j = 0, \ldots, p-1)$, a truncated $N(u_\alpha, 1/v_\alpha)$ for $\alpha_i$ with $\alpha_i \in (-1, 1)$ and $\text{Gamma}(u_\tau, v_\tau)$ for $\tau_i$. For the changepoint $m$, we use a discrete Uniform prior on $\{1, \ldots, n-1\}$, denoted $DU[1, n-1]$. To complete the prior specification, let $\pi(\mathcal{M}_0) = \pi(\mathcal{M}_{00}) = 1/2$, where $\mathcal{M}_0$ and $\mathcal{M}_{00}$ are the model indicators for the no-changepoint and the changepoint models with parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}$, respectively.

Letting $\boldsymbol{\theta}_{-\phi}$ denote the vector of parameters excluding a parameter $\phi$ and $n_i = \sum_{t=1}^{n} I_{\{c_t = i\}}$ $(i = 1, 2)$, the conditional distributions of the parameters are:

$$\boldsymbol{\beta}_i | \boldsymbol{\theta}_{-\boldsymbol{\beta}_i}, \mathbf{x}, \mathbf{y}, \mathbf{z} \propto \exp\left( \sum_{t=1}^{n} I_{\{c_t = i\}}(\mathbf{z}_t^T \boldsymbol{\beta}_i y_t - e^{\mathbf{z}_t^T \boldsymbol{\beta}_i + x_t}) - \frac{v_\beta}{2} \sum_{j=0}^{p-1}(\beta_{i,j} - u_\beta)^2 \right);$$

$$\tag{5}$$

$$\alpha_i | \boldsymbol{\theta}_{-\alpha_i}, \mathbf{x}, \mathbf{y} \propto (\sqrt{1 - \alpha_i^2})^{I_{\{i=1\}}} \exp\left( -\frac{A_i}{2} \left\{ \alpha_i - \frac{B_i}{A_i} \right\}^2 \right) I(|\alpha_i| < 1); \tag{6}$$

$$\tau_i | \boldsymbol{\theta}_{-\tau_i}, \mathbf{x}, \mathbf{y} \sim \text{Gamma}\left( u_\tau + \frac{n_i}{2}, \frac{C_i}{2} + v_\tau \right); \tag{7}$$

$$m | \boldsymbol{\theta}_{-m}, \mathbf{x}, \mathbf{y}, \mathbf{z} = \frac{p(\boldsymbol{\theta}_{-m}, m; \mathbf{x}, \mathbf{y}, \mathbf{z})}{\sum_{j=1}^{n-1} p(\boldsymbol{\theta}_{-j}, j; \mathbf{x}, \mathbf{y}, \mathbf{z})}; \tag{8}$$

where

$$p(\boldsymbol{\theta}_{-m}, m; \mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \prod_{i=1}^{2} \left\{ \exp\left( \sum_{t=1}^{n} I_{\{c_t=i\}}(\mathbf{z}_t^T \boldsymbol{\beta}_i y_t + x_t y_t - \exp(\mathbf{z}_t^T \boldsymbol{\beta}_i + x_t)) \right) \right.$$
$$\left. \times \tau_i^{\frac{n_i}{2}} \exp\left( -\frac{\tau_i}{2} \sum_{t=2}^{n} I_{\{c_t=i\}}(x_t - \alpha_i x_{t-1})^2 \right) \right\},$$

$A_i = \tau_i \sum_{t=2}^{n} I_{\{c_t=i\}} x_{t-1}^2 + v_\alpha$, $B_i = \tau_i \sum_{t=2}^{n} I_{\{c_t=i\}} x_t x_{t-1} + u_\alpha v_\alpha$ and $C_i = \sum_{t=2}^{n} I_{\{c_t=i\}}(x_t - \alpha_i x_{t-1})^2 + I_{\{i=1\}}(1 - \alpha_1^2)x_1^2$. The above equations hold for the model without a changepoint with $\boldsymbol{\theta}_2$ being vacuous. Note that the conditional posterior distribution of $\alpha_2$ is Normal. By the Markov property due to the AR(1) assumption, $p(x_t|\mathbf{x}_{-t}, \boldsymbol{\theta}) \propto p(x_t|x_{t-1}) \times p(x_{t+1}|x_t)$, the conditional posterior distribution of $x_t$ for $t = 2, \ldots, n-1$ can easily be derived from equation (4) as

$$\pi(x_t|\mathbf{x}_{-t}, \mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) \propto \exp\left( -\frac{\tau_{c_t}}{2}(x_t - \alpha_{c_t} x_{t-1})^2 \right) \times \exp\left( -\frac{\tau_{c_{t+1}}}{2}(x_{t+1} - \alpha_{c_{t+1}} x_t)^2 \right)$$
$$\times \exp\left( x_t y_t - e^{\mathbf{z}_t' \boldsymbol{\beta}_{c_t} + x_t} \right). \tag{9}$$

We give details of how the above conditional distributions are used in the MCMC algorithm in Section 2.2 below.

## 2.2 The reversible jump MCMC algorithm

In proposing the single changepoint model, our goal is not limited to making inference about the parameters of the model. We are also interested in incorporating within our Bayesian framework, a procedure for testing for the presence or otherwise of a changepoint in the data. In this Section, we provide the details of a RJMCMC algorithm for testing the presence of a single changepoint and highlight how this can be extended to multiple changepoints in Section 5 and the supplementary material. The algorithm alternates between two updating steps: the within-model updates and the model-switching step.

**Within-model updates**. This updating scheme involves sampling the parameters of the current model using standard Metropolis-Hastings updates. For $\alpha_2$, $\tau_1$ and $\tau_2$ full conditional distributions are in a standard form and there-

fore a Gibbs step can be used to update these parameters. For $\alpha_1$, we use an independence sampler proposing from a $N(B_1/A_1, 1/A_1)$, where $A_1$ and $B_1$ are defined after (8). This is a good approximation for the conditional distribution of $\alpha_1$ and we only need to correct for $\sqrt{1 - \alpha_1^2}$ in the acceptance probability. For $i = 1, 2$, the conditional distribution of $\boldsymbol{\beta}_i$ does not have a nice form and therefore we update this parameter block using a random walk Metropolis step. For updating $m$, we found that a hybrid proposal scheme worked well where with probability $p$ we propose a local move and otherwise made a global move. For the local move we propose to move the changepoint to $m + VQ$, where $V$ takes the values $\pm 1$ with probability 0.5 and $Q \sim \text{Geometric}(q)$ with support on $\mathbb{N}$. For the global move, the proposed changepoint is drawn uniformly at random from $\{1, 2, \ldots, n - 1\}$. Over a range of examples, we found that $p = 0.9$ and $q = 0.5$ worked well.

The augmented data (latent process) are updated using a proposal scheme relative to independence sampling procedure. We exploit the AR(1) structure of the latent process to construct a suitable proposal distribution. For $x_2 \ldots, x_m$, and $x_{m+2} \ldots, x_{n-1}$ the proposal distribution is

$$
\begin{aligned}
q(x_t | \mathbf{x}_{-t}, \boldsymbol{\theta}) &\propto \exp\left(-\frac{\tau_{c_t}}{2}(x_t - \alpha_{c_t} x_{t-1})^2\right) \times \exp\left(-\frac{\tau_{c_t}}{2}(x_{t+1} - \alpha_{c_t} x_t)^2\right) \\
&\sim N\left(\frac{\alpha_{c_t}(x_{t-1} + x_{t+1})}{1 + \alpha_{c_t}^2}, \frac{1}{\tau_{c_t}(1 + \alpha_{c_t}^2)}\right), \quad t = 2, \ldots, n-1, \quad (10)
\end{aligned}
$$

where $\mathbf{x}_{-t}$ denotes the vector of observations of the latent process excluding the $t^{\text{th}}$ value, and the corresponding acceptance probability is

$$
\alpha(x_t, x_t') = \min\left\{1, \frac{\exp(x_t' y_t - e^{x_t' + \mathbf{z}_t^T \boldsymbol{\beta}_{c_t}})}{\exp(x_t y_t - e^{x_t + \mathbf{z}_t^T \boldsymbol{\beta}_{c_t}})}\right\}, \quad (11)
$$

where $x_t'$ denotes the proposed value of $x_t$. Clearly, the proposal distribution of $x_t'$ is motivated by the Markovian configuration of the latent process which is then *corrected* to take into account $y_t$. The proposal distributions for $x_1$, $x_{m+1}$ and $x_n$ can be similarly defined taking into account the special nature of these three quantities.

**Model-switching step.** This updating step of the RJMCMC sampler involves proposing transitions between the competing models. Parameter estimation for model $\mathcal{M}_0$ can be done using the above procedures excluding the steps for updating $\boldsymbol{\theta}_2$ and $m$. Since model $\mathcal{M}_{00}$ has more parameters than model $\mathcal{M}_0$ it is necessary to introduce auxiliary parameters to move from model $\mathcal{M}_0$ to model $\mathcal{M}_{00}$, see Green (1995). The auxiliary parameters are: $\tilde{a}_i (i = 0, \dots, p - 1)$, $\tilde{b}$ and $\tilde{c}$, which are drawn from $N(0, \sigma_a^2)$, $N(0, \sigma_b^2)$ and $N(0, \sigma_c^2)$, respectively; $\sigma_a^2$, $\sigma_b^2$ and $\sigma_c^2$ are typically pilot-tuned to achieve efficient moves. With $k$, we choose a position for the proposed changepoint uniformly at random from the interval, that is, $k \sim DU[1, n - 1]$. The $\tilde{a}_i$'s are needed to compensate for the absence of the second set of regression coefficients in $\mathcal{M}_0$ while $\tilde{b}$ and $\tilde{c}$ compensate for the absence of a second $\alpha$ and $\tau$ term, respectively. We then propose to move to model $\mathcal{M}_{00}$ with model parameters $\boldsymbol{\theta}$ with a changepoint at $m = k$, $(\boldsymbol{\beta}_1', \alpha_1', \tau_1') = (\boldsymbol{\beta}_1, \alpha_1, \tau_1) + (1 - k/n)(\tilde{\boldsymbol{a}}, \tilde{b}, \tilde{c})$ and $(\boldsymbol{\beta}_2', \alpha_2', \tau_2') = (\boldsymbol{\beta}_1, \alpha_1, \tau_1) - k/n(\tilde{\boldsymbol{a}}, \tilde{b}, \tilde{c})$. Note that if we propose a negative value for $\tau_i$ $(i = 1, 2)$ then the proposed move is rejected. For the reverse move, no auxiliary parameters are needed. Consequently, we propose the parameters of model $\mathcal{M}_0$ as the weighted means of the corresponding parameters of model $\mathcal{M}_{00}$ with $\alpha_1' = (m/n)\alpha_1 + (1 - (m/n))\alpha_2$, $\boldsymbol{\beta}_1' = (m/n)\boldsymbol{\beta}_1 + (1 - (m/n))\boldsymbol{\beta}_2$ and $\tau_1' = (m/n)\tau_1 + (1 - (m/n))\tau_2$. The above proposal scheme is based on the idea of moment-matching, see Green (1995) and Brooks et al. (2003) and is found to work effectively here.

Observe that the move from $\mathcal{M}_0$ to $\mathcal{M}_{00}$ increases the dimension of the parameter vector by a factor of $p + 3$. Consequently, we found on the basis of a simulation study that the algorithm exhibited an undue preference for $\mathcal{M}_{00}$, especially as $n$ increases. Similar problems were observed for reversible jump MCMC algorithms for $INARMA$ models in Enciso-Mora et al. (2009a) and Enciso-Mora et al. (2009b). The solution proposed in Enciso-Mora et al. (2009a) and Enciso-Mora et al. (2009b) was to use a 'BIC-like' penalization with the prior on models $\mathcal{M}_0$ and $\mathcal{M}_{00}$ satisfying $\pi(\mathcal{M}_0) \propto n^{-\frac{1}{2}\dim(\boldsymbol{\theta}_1)}$ and $\pi(\mathcal{M}_{00}) \propto n^{-\frac{1}{2}\dim(\boldsymbol{\theta})}$. We employed the 'BIC-like' penalization for the model and found that it gave consistent estimation of the posterior probability of the models

9

as $n$ was varied. Given that at each iteration of the algorithm we propose to switch model, the acceptance probability for moving from $\mathcal{M}_0$ to $\mathcal{M}_{00}$ is given by $\alpha_{0,00} = \min\{1, A_{0,00}(\boldsymbol{\theta}_1, \boldsymbol{\theta})\}$. The acceptance ratio $A_{0,00}(\boldsymbol{\theta}_1, \boldsymbol{\theta})$ is given by

$$A_{0,00}(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = \frac{\pi(\mathbf{x}, \mathbf{y}, \mathbf{z} | \boldsymbol{\theta}, \mathcal{M}_{00})}{\pi(\mathbf{x}, \mathbf{y}, \mathbf{z} | \boldsymbol{\theta}_1, \mathcal{M}_0)} \times \frac{\pi(\mathcal{M}_{00}, \boldsymbol{\theta})}{\pi(\mathcal{M}_0, \boldsymbol{\theta}_1)} \times \frac{q(\boldsymbol{\theta}, \boldsymbol{\theta}_1)}{q(\boldsymbol{\theta}_1, \boldsymbol{\theta})} \times |J_{0 \to 00}|, \qquad (12)$$

where $\frac{\pi(\mathbf{x}, \mathbf{y}, \mathbf{z} | \boldsymbol{\theta}, \mathcal{M}_{00})}{\pi(\mathbf{x}, \mathbf{y}, \mathbf{z} | \boldsymbol{\theta}_1, \mathcal{M}_0)}$ denotes the ratio of the likelihoods for $\mathcal{M}_{00}$ and $\mathcal{M}_0$, $\frac{\pi(\mathcal{M}_{00}, \boldsymbol{\theta})}{\pi(\mathcal{M}_0, \boldsymbol{\theta})}$ is the corresponding ratio of the prior distributions and $\frac{q(\boldsymbol{\theta}, \boldsymbol{\theta}_1)}{q(\boldsymbol{\theta}_1, \boldsymbol{\theta})}$ is the ratio of the proposal distributions for the move from $\mathcal{M}_{00}$ to $\mathcal{M}_0$ and the reverse move from $\mathcal{M}_0$ to $\mathcal{M}_{00}$. Note that $q(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = 1$ since the move from $\mathcal{M}_{00}$ to $\mathcal{M}_0$ is deterministic. Finally, $J_{0 \to 00}$ is the Jacobian for the transformation between the parameters of the two models. Due to the block diagonal nature of $J_{0 \to 00}$ it is straightforward to show that $|J_{0 \to 00}| = 1$. Explicit derivations of the terms in (12) are given in the supplemental material for this paper. Note that the acceptance ratio for the reverse move (i.e. $\mathcal{M}_{00} \to \mathcal{M}_0$) is simply obtained using the inverse of all the terms in equation (12).

# 3 Simulation study

In order to assess the performance of the changepoint model and the RJMCMC algorithm we conducted a simulation study. The simulation study was designed to check whether or not we could correctly identify the location of a changepoint in the model.

The simulation study was as follows. We constructed a covariate vector $\mathbf{z}^T = (1, \sin(2\pi t/6), \cos(2\pi t/6), N(0,1))$ and generated 9 data sets each with $n = 800$ and $m = 400$. For all 9 data sets $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_{1(1\times4)}^T, \alpha_1, \tau_1) = (0.1, 0.2, 0.1, 0.3, 0.1, 4.0)$, whilst $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1 + c$, where $c = 0, 0.1, \ldots, 0.8$, in successive data sets. Therefore in data set 1 there is effectively no changepoint, whereas for data set 9 there is a very emphatic changepoint in the parameters, especially $\alpha$. The prior distributions used are: $\beta_{i,j} \sim N(0,1)$, $\alpha_i \sim N(0.5, 1/0.5) \, I(|\alpha_i| < 1)$, $\tau_i \sim \text{Gamma}(3,1)$, $m \sim DU[1, n-1]$, for $i = 1, 2$ and $j = 0, \ldots, p-1$. For each data set,

the algorithm was applied successively to $\mathbf{x}_{(l+1):(l+400)} = (x_{l+1}, \ldots, x_{l+400})$ for $l = 0, 10, \ldots, 400$ to estimate the changepoint and the model parameters. Thus the algorithm was applied to 41 subsets of length 400 of each of the 9 datasets in which, $l = 0$ and $l = 400$ correspond to no changepoint, whilst $l = 200$ corresponds to a changepoint in the middle of the series. Each time, our RJMCMC sampler was initialized in model $\mathcal{M}_0$ (except for $l = 0$ and $l = 400$ where we initialized in model $\mathcal{M}_{00}$) and run for 45,000 iterations following a burn-in period of 5,000 iterations. Visual examination of the trace plots of the parameters showed that convergence had been attained before the end of the burn-in period, with in most cases all the parameters being well estimated. Some trace plots of the changepoint $m$ and other parameters are included in the supplemental material.

[Figure 1 about here]

Figure 1 shows the posterior probabilities of the changepoint model for $c = 0.0, 0.1, \ldots, 0.5$. The posterior probability of a changepoint increases with $c$ up to $c = 0.5$ and for $c > 0.5$, the posterior probability of a changepoint is close to one for all segments except the first segment (no changepoint with parameters $\boldsymbol{\theta}_1$). We note that until segment 5 or 6 where there are 40 or 50 observations with $\boldsymbol{\theta}_2$ parameters the posterior probability of a changepoint is low whereas for the latter segments where most of the observations are with the $\boldsymbol{\theta}_2$ parameters the posterior probability of a changepoint drops off more slowly. For segments 6 to 36 where there are at least 50 observations from each set of parameters consistent estimation of the changepoint is observed for $c \geq 0.3$ with the posterior mode of the changepoint close to the true changepoint and the 95% credible interval for the changepoint typically having a small range of between 25 and 50. For segments 37 to 41 where almost all observations come from the $\boldsymbol{\theta}_2$ parameters we observe poor estimation of the location of the changepoint, especially when $c$ is large with significant correlation in the latent process ($\alpha_1 \geq 0.5$). In such cases the 95% credible interval for the changepoint covers most of the range with the mode for the changepoint close to the end rather than the beginning of the segment analysed.

# 4 Data analysis

## 4.1 Benefit claims data

This data comprises of monthly counts of claimants collecting wage loss benefit for injuries in the workplace from January 1985 to December 1994. It has been previously analyzed in Freeland & McCabe (2004) and Enciso-Mora et al. (2009a). Figure 2 shows the plot of the data. Its mean and variance are 6.13 and 11.80 respectively; which indicate overdispersion.

[Figure 2 about here]

A visual examination of the plot and an exploratory analysis indicate a change in mean and variance occurring around the beginning of 1990. The data also shows evidence of seasonality. This seeming change in the behaviour of the series was represented/interpreted in previous analyses as a decrease in trend (see, for example, Enciso-Mora et al., 2009a, and $\beta_1$ in Table 2). However, it can be seen from Figure 2 that the data do not exhibit any trend if we presuppose that a change has occurred around 1990. Further exploratory analysis using an ordinary Poisson regression model (with independent outcomes) fitted in a non-Bayesian framework also confirms this. We, thus, excluded trend from the changepoint analysis. (The inclusion of trend resulted in poor estimation of the changepoint parameter. We compare the Bayesian changepoint analysis without trend and a Bayesian analysis with trend but no changepoint in the discussion below.) We use the covariate vector: $\mathbf{z}_t^T = (1, \sin(2\pi t/6), \cos(2\pi t/6))$, with the sine and cosine terms at the semi-annual frequency characterizing seasonality in the model. The prior distributions used are: $N(0.2, 1/0.2)$ for $\alpha_1, \alpha_2 \in (-1, 1)$, Gamma(5, 1) for $\tau_1, \tau_2$ and $N(0, 1)$ for the regression coefficients.

The RJMCMC algorithm was run for 1,000,000 iterations, with the first 10% of the iterations discarded as burn-in. Some of the results of the analysis are plotted in Figure 3. The model indicator in part (a) of the figure shows that the RJMCMC algorithm mixed well, spending more time in the changepoint model which has a posterior probability of 0.9658. The changepoint $m$ has a posterior mode of 55 which corresponds to July, 1989. Its 95% credible interval is [44,

66] = [August 1988, June 1990]. Hence, even though a changepoint seems to have occurred around the beginning of 1990 by visual examination, our analysis shows that this is not exactly the case. The unimodal posterior distribution of $m$ in part (b) of the figure provides evidence of a single changepoint. (The trace plot of $m$ is included in the supplementary material.) Also, when the algorithm was applied to the data before and after the changepoint, no other changepoints were detected.

[Table 1 about here]

The estimates in Table 1 show a slight decrease in the lag-1 autocorrelation of the latent process and its variance as well as the regression coefficients after the changepoint. Thus accounting for the general decrease in the number of claims after this time. This decrease in the number of claims may be due to some unreported changes in working conditions or insurance policies. Figure 3 (c) shows that the fitted values (longdashed line) provide good estimates of the observed values. The partial autocorrelation function (pacf) plot of the residuals in part (d) of the figure shows that the residuals are an approximate white noise process. It can thus be inferred that the model fits the data reasonably well. By identifying significant autocorrelation in the latent process (as evidenced by the estimates of $\alpha_1$ and $\alpha_2$) both pre-and post-change, our analysis agrees with those of Freeland & McCabe (2004) and Enciso-Mora et al. (2009a).

[Figure 3 about here]

As we noted above, previous analyses of this data using no-changepoint models included trend in the covariate vector. We now compare this approach with our changepoint analysis. We refer to the no-changepoint model with trend in the covariate vector as model $M_1$ (see equation (1)) and the changepoint model without trend as model $M_2$. Thus, for $M_1$, the covariate vector is $\mathbf{z}_t^T = (1, t/n, \sin(2\pi t/6), \cos(2\pi t/6))$. Using the same conditions as before, the parameter estimates of $M_1$ are shown in Table 2.

[Table 2 about here]

These parameter estimates are reasonably similar to those reported in Table 1. The fitted values of $M_1$ also plotted in Figure 3 (c) (dotted line) show that both models produce very close estimates of the data. As additional crite-

13

ria to compare both models, we employed the root mean square error $\left(\text{RMSE}\right.$ $=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}$, where $\hat{y}_t$ is the predicted value of $y_t\big)$ and the predictive model choice criterion $\big(\text{PMCC}=\sum_{t=1}^{n}(y_t-E(y_t^{\text{rep}}))^2+\sum_{t=1}^{n}Var(y_t^{rep})$, where $y_t^{\text{rep}}$ denotes the predictive replicate of $y_t\big)$ proposed by Gelfand & Ghosh (1998). The first term of the PMCC provides a measure of goodness-of-fit while the second term is a penalty term which increases with model complexity. For both criteria, the smaller the value, the better the fit. Note that other suitable model choice criteria could also be used. From our analyses, the changepoint model gave an RMSE value of 1.55 while that of the no-changepoint model with trend was 1.64. Using the PMCC, we obtained a value of $286.53+1055.42 = 1341.95$ for the changepoint model without trend, whereas with the no-changepoint model with trend, we obtained $324.62 + 1023.26 = 1347.88$. This shows that even though the changepoint model uses more parameters to model the data, it provides a better fit according to these criteria. Our changepoint analysis thus offers a new perspective to understanding the dynamics of the data.

## 4.2 Presidential uses of force in the US

As a second example, we applied the model to the annual counts of instances in which US presidents used force abroad. The data were downloaded from http://bingweb.binghamton.edu/~bfordham. The details of these events and other information about the data can also be found at the website. Although the original data spans between 1870 and 1995, the sample period for this analysis is restricted to cover from 1930 to 1995 since covariate data are readily available for this period. Previous political studies (Fordham, 1998; Gowa, 1998; Park, 2006) have shown that the condition of the domestic economy plays a significant role in presidential decisions to use force abroad. Inidicators of the domestic economy available for this analysis were the Gross Domestic Product (GDP) and the Unemployment rates (UNEMP, given in percentages) for the time period studied. However, exploratory analysis showed that GDP did not have a significant relationship with use of force. Hence, we included only UN-EMP as a covariate in our analysis.

[Figure 4 about here]

Figure 4 shows that there is evidence of an increasing trend in use of force; thus, we also included trend in our covariate vector. This increasing trend is also confirmed by an exploratory non-Bayesian analysis using an ordinary Poisson regression model. The mean and variance of the data are 6.15 and 24.04 respectively; both of which suggest overdispersion. A closer examination of the plot also reveals that there appears to be a change in this increasing trend around 1965. We use the following prior distributions for our analysis: $N(0.5, 1/0.5)$, Gamma $(5, 1)$ and $N(0, 1)$ for $\alpha_i \in (-1, 1)$, $\tau_i$ and $\beta_{i,j}(i = 1, 2; \quad j = 0, \ldots, p-1)$. The covariate UNEMP was standardized to aid the convergence of our algorithm.

[Table 3 about here]

The reversible jump MCMC algorithm was run for 1,000,000 iterations. We used the last 900,000 iterations for our analysis, discarding the first 100,000 as burn-in. The trace plot of the model indicator is shown in part (a) of Figure 5.

[Figure 5 about here]

This clearly shows that the algorithm mixed well. The posterior probability of the changepoint model is 0.9529 which provides a very strong indication of the presence of a changepoint. The posterior mode of the changepoint is $\hat{m} = 37$ which corresponds to the year 1966 while its 95% credible interval is [23, 54]=[1952, 1983]. The posterior distribution of the changepoint $m$ in part (b) of Figure 5 evidences the presence of a single changepoint in the data. (The trace plot of $m$ is included in the supplementary material.) Also, an application of the algorithm to the subsets of the data before and after 1966 revealed that there was no other changepoint. Figure 5 (c) shows a sharp increase in Presidential use of force before 1966 which mostly occurred during the Kennedy-Johnson era (1958-1966). This period is followed by a slight decrease in the number of instances of use of force abroad which has been partly attributed to the United States' increased commitment to South Vietnam during the Vietnam war (see Park (2006)).

15

Table 3 reports the differences in the estimates of the parameters before and after the changepoint. It can be seen that there are significant changes in the values of the parameters. In particular, we see a dramatic increase in the (positive) effect of UNEMP on use of force after the changepoint. This effect can be explained by the fact that high UNEMP reduces labour cost thereby making use of force less costly and more attractive (see Fordham, 1998). Figure 5(c) plots the observed data and the posterior means of the fitted values. This shows that the model yields reasonable estimates of the data. The pacf plot in part (d) of the figure shows that the residuals mimic a white noise process. Hence, the model fits the data reasonably well.

We note that previous analyses of this data such as Fordham (1998) and Park (2006) used log-linear models which did not account for serial autocorrelation. In contrast, our analysis shows that there exists significant autocorrelation in the data as can be seen in the values of $\alpha_1$ and $\alpha_2$ in Table 3.

# 5  Discussion

We have developed a methodology for detecting and estimating a single changepoint in a parameter-driven model. In our model formulation, a changepoint is characterized by a change in the parameters of the latent process as well as the regression coefficients. Although this changepoint parameterization presents a general picture of the problem, other changepoint parameterizations could be explored as we have noted in Section 2. The proposed RJMCMC algorithm has been shown to perform well with both simulated and real life data sets. It should be noted that the performance of the algorithm has been tested in the presence of covariates as these are important and help to explain the observed data in the regression context. In addition, the covariates included in the model also help to characterise the changepoint as can be seen in equation (3). In the examples in Section 4, different results may be obtained in the absence of covariates or with different sets of covariates (see, for example, Park, 2006).

There are limitations to assuming a single changepoint. The aim of this paper was to present a framework for detecting a changepoint in count data time

series models. In the applications in Section 4, we have affirmed the presence of a single changepoint by applying the algorithm to the segments created by the changepoint. For datasets with multiple changepoints, the application of the single changepoint algorithm will most likely lead to a multimodal distribution for the changepoint parameter, $m$. And upon segmenting the data using the highest mode and applying the algorithm successively to the segments, the other changepoints could be detected. This procedure is, however, ad hoc and will not be practical in many scenarios. Hence, a natural extension of the methodology is to construct a RJMCMC algorithm which detects the location(s) of an unknown number of multiple changepoints. This can be done relatively straightforwardly by treating subsets of the time series as time series with either 0 or 1 changepoint and seeking to add or remove a changepoint from the subset. Thus the single changepoint methodology developed here forms the building block for a more extensive changepoint analysis and this approach will be based on the changepoint RJMCMC introduced in Green (1995), Section 3.

An interesting alternative to the changepoint modelling is to assume that there is a multi-state hidden Markov model for the latent process, see Chib (1998). This is appropriate when the model parameters move between a finite set of parameters rather than a change leading to a completely new set of parameter values.

Lastly, although the single Metropolis-Hastings update proposed in this paper for updating the latent process has been shown to perform well, there could be further gains in terms of mixing and speed of convergence of the algorithm when a block updating procedure such as particle filtering MCMC is used (see, e.g., Whiteley et al., 2009).

## Supporting Information

The Supporting Information for this paper contains explicit derivations of the terms in equation (12) and some trace plots from Sections 3 and 4. We also discuss the extension to multiple changepoints in detail.

# Acknowledgements

# References

Barbieri, M. M. & Conigliani, C. (1998). Bayesian analysis of autoregressive time series with change points. *Journal of the Italian Statistical Society*, 3, 243–255.

Brooks, S. P., Guidici, P., & Roberts, G. O. (2003). Efficient construction of reversible jump Markov Chain Monte Carlo proposal distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(1), 3–55.

Cameron, A. C. & Trivedi, P. K. (2013). *Regression analysis of count data.* New York: Cambridge University Press, second edition edition.

Campbell, M. J. (1994). Time series regression for counts: An investiagtion into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society A*, 157, 191–208.

Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical bayesian analysis of changepoint problems. *Applied Statistics*, 41, 389–405.

Chan, K. S. & Ledolter, J. (1995). Monte carlo EM estimation for time series model involving counts. *J. Amer. Statist. Assoc.*, 90(429), 242–252.

Chib, S. (1998). Estimation and comparison of multiple changepoint models. *J. Econometrics*, 86, 221–241.

Davis, R. A., Dunsmuir, W. T., & Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, 87(3), 491–505.

Enciso-Mora, V., Neal, P., & Subba-Rao, T. (2009a). Efficient order selection algorithms for integer-valued ARMA processes. *J. Time Series Anal.*, 30(1), 1–18.

Enciso-Mora, V., Neal, P., & Subba Rao, T. (2009b). Integer-valued ar processes with explanatory variables. *Indian Journal of Statistics*, 71(B), 248–263.

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple change-point problems. *Statist. Comput.*, 16(2), 203–213.

Fordham, B. (1998). The politics of threat perception and the use of force: A political economy model os us uses of force, 1949-1994. *International Studies Quarterly*, 42(3), 567–590.

Freeland, R. K. & McCabe, B. P. M. (2004). Analysis of low count time series data by Poisson autoregression. *J. Time Series Anal.*, 25(5), 701–722.

Frühwirth-Schnatter, S. & Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. Biometrika, 93(4), 827–841.

Fu, Y. & Curnow, R. N. (1990). Maximum likelihood estimation of multiple change points. *Biometrika*, 77(3), 563–73.

Gelfand, A. E. & Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85, 1–11.

Gowa, J. (1998). Politics at the water's edge: Parties, voters, and the use of force abroad. *International Organisation*, 52(2), 307–324.

Green, P. J. (1995). Reversible jump Markov Chain Monte Carlo computation and bayesian model determination. *Biometrika*, 82(4), 711–32.

Jandhyala, V., Fotopoulos, S., MacNeill, I., & Liu, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.*, (Early View), doi: 10.1111/jtsa12035.

Kang, J. & Lee, S. (2014). Parameter change test for poisson autoregressive models. *Scand. J. Stat.*, 41, 1136–1156.

Oh, M. & Lim, Y. B. (2001). Bayesian analysis of time series Poisson data. *J. Appl. Stat.*, 28(2), 259–271.

Park, J. H. (2006). Modeling structural changes: Bayesian estimation of multiple changepoint models and state-space models. *Prepared for American Political Science Association Meeting, August, 2006, Philadelphia.*

Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.*, 82(398), 528–540.

Utazi, C. E. (2014). *Spatio-Temporal Modelling of Partially Observed Processes.* PhD thesis, Lancaster University.

Whiteley, N., Andrieu, C., & Doucet, A. (2009). Particle Markov Chain Monte Carlo for multiple change-point problems. *Technical report no. 0911 Department of Mathematics Bristol University*, (pp. 1–15).

Zeger, S. L. (1988). A regression model for times series of counts. *Biometrika*, 75(4), 621–9.

Chigozie E. Utazi, Southampton Statistical Sciences Research Institute and WorldPop, Department of Geography, University of Southampton, UK, SO17 1BJ.
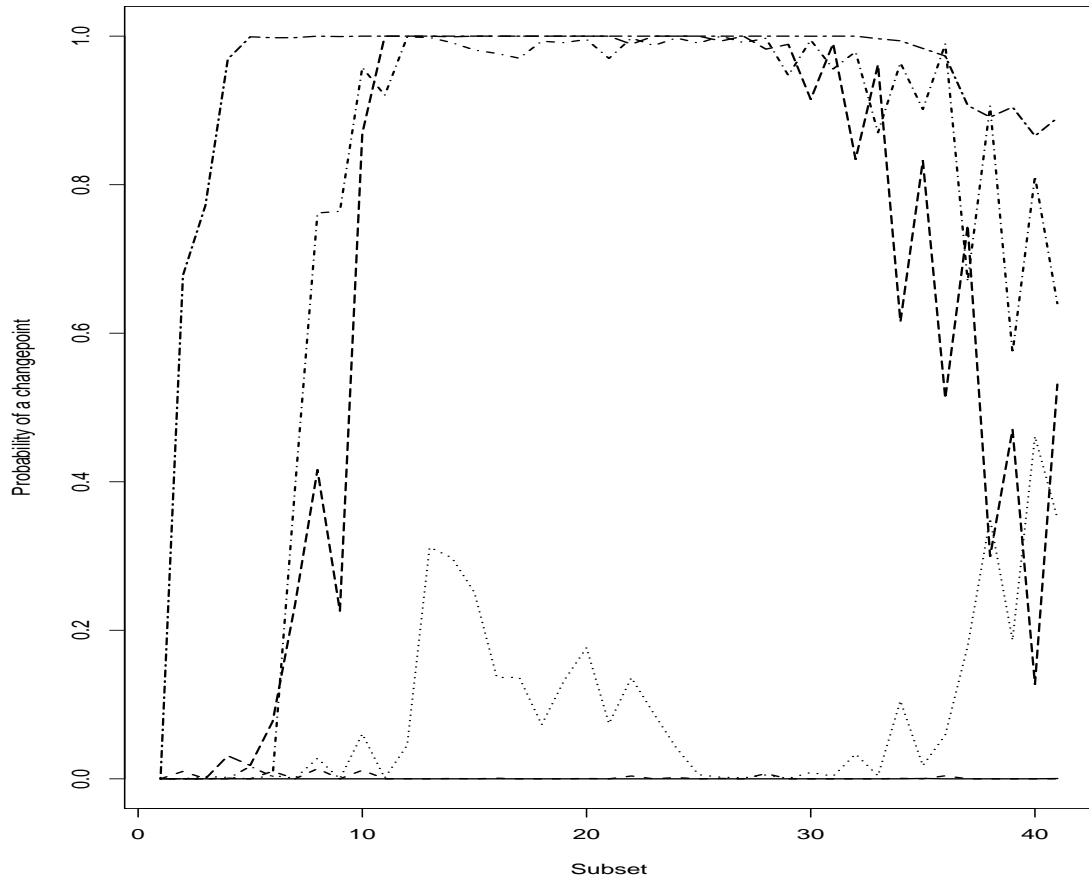
Email: c.e.utazi@soton.ac.uk

Figure 1: Plots of the posterior probabilities of the changepoint model; $c = 0.0$ - solid, $c = 0.1$ - dashed, $c = 0.2$ - dotted, $c = 0.3$ - dotdash, $c = 0.4$ - longdash, $c = 0.5$ - twodash.
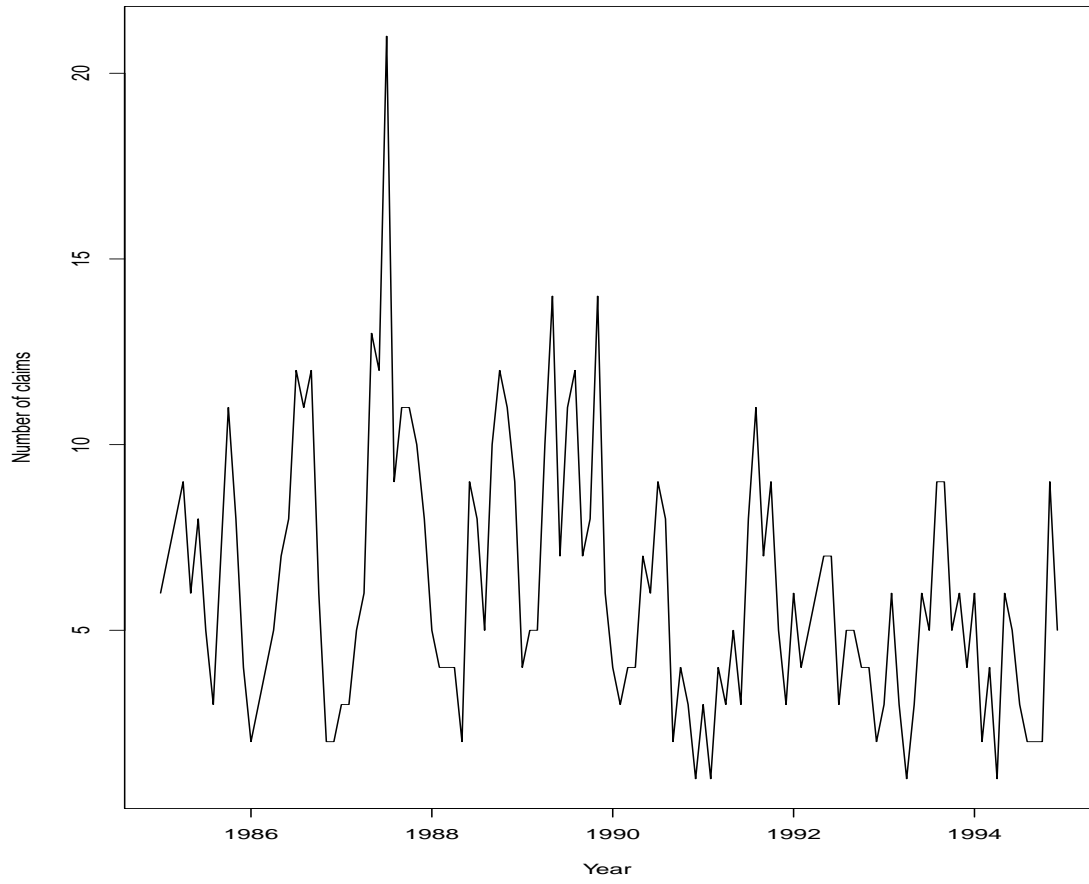
Figure 2: Time series plot of the benefit claims data from January 1985 to December 1994.

Table 1: Posterior means and standard errors (s.e.) of the parameters for the benefit claims data

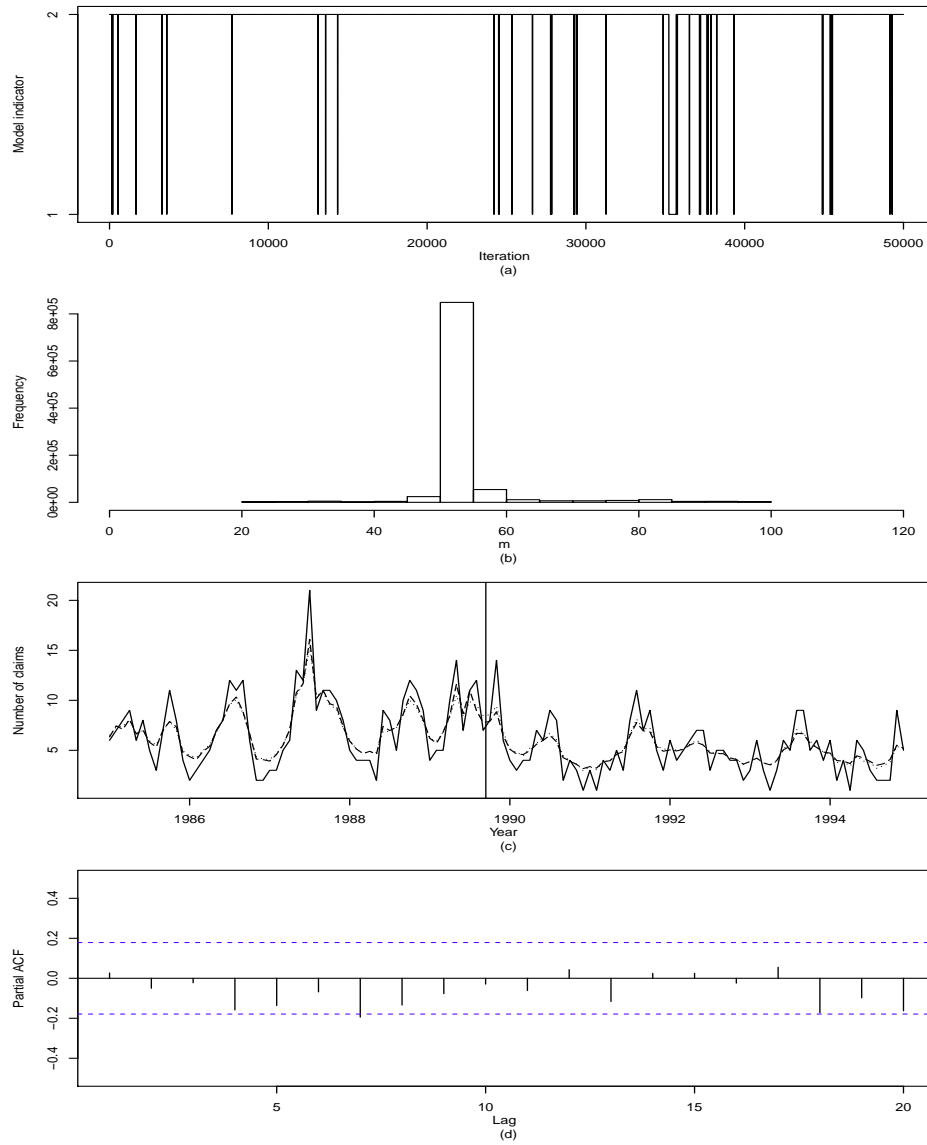| Parameter | Mean (s.e.) | Parameter | Mean (s.e.) |
|---|---|---|---|
| $\alpha_1$ | 0.6343 (0.1770) | $\alpha_2$ | 0.5010 (0.2552) |
| $\tau_1$ | 7.3217 (1.6689) | $\tau_2$ | 8.0307 (1.9937) |
| $\beta_{0,1}$ | 1.7433 (0.2570) | $\beta_{0,2}$ | 1.4786 (0.2373) |
| $\beta_{1,1}$ | 0.0285 (0.0947) | $\beta_{1,2}$ | 0.0114 (0.1567) |
| $\beta_{2,1}$ | 0.0515 (0.0925) | $\beta_{2,2}$ | 0.0065 (0.1860) |

Figure 3: Benefit claims data. (a) Trace plot of the model indicator for the last 50,000 iterations: 1 = no-changepoint model and 2 = changepoint model. (b) Histogram of the posterior distribution of $m$. (c) Observed data (solid line) vs posterior means of the fitted values (longdashed line). The vertical line indicates the mode of the changepoint. The dotted line shows the fitted values of the no-changepoint model with trend (d) Partial autocorrelation plot of the residuals. The dotted lines indicate the $\pm 2$ standard error confidence limits.

Table 2: Posterior means and standard errors (s.e.) of the parameters of the no-changepoint model with trend

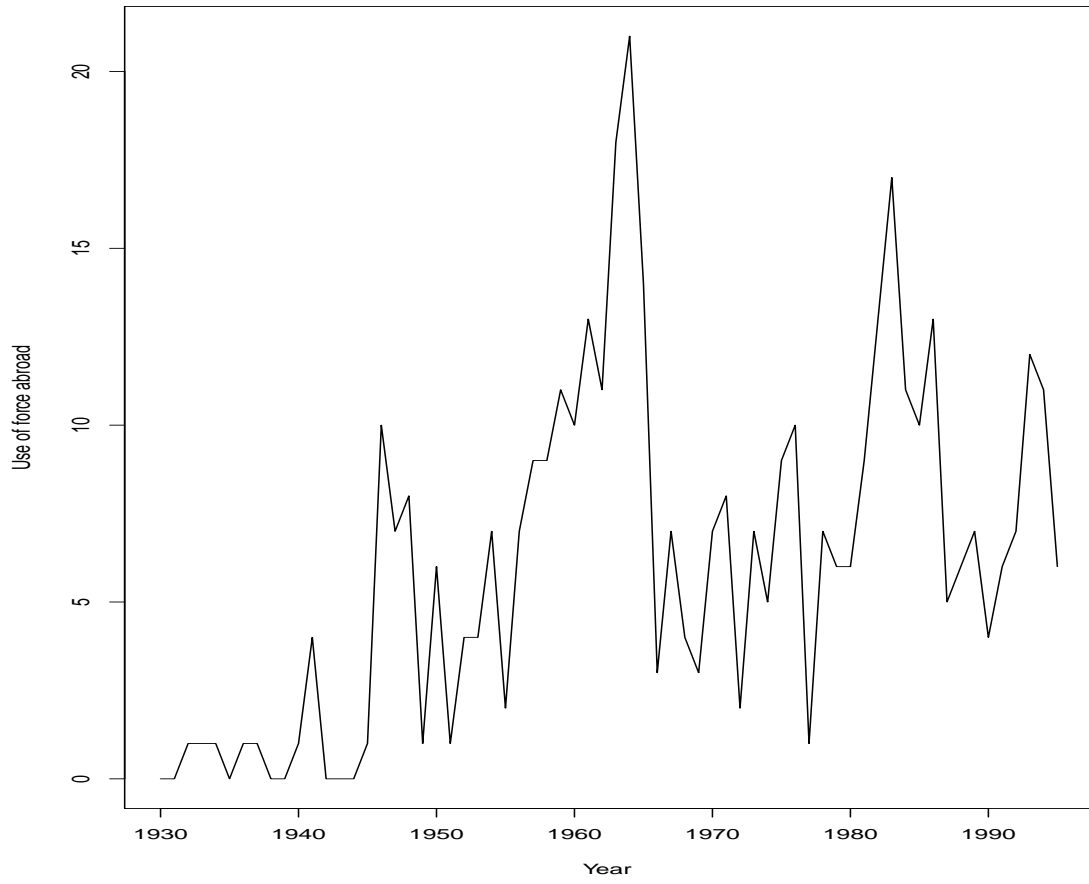| Parameter | $\alpha$ | $\tau$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
|-----------|----------|--------|-----------|-----------|-----------|-----------|
| Mean | 0.7167 | 13.8353 | 1.9475 | -0.4722 | 0.0276 | 0.0356 |
| (s.e.) | (0.1029) | (4.7623) | (0.2506) | (0.2956) | (0.0576) | (0.0568) |

Figure 4: Presidential uses of force in the US between 1930 and 1995

Table 3: Posterior means and standard errors (s.e.) of the parameters for the presidential uses of force data

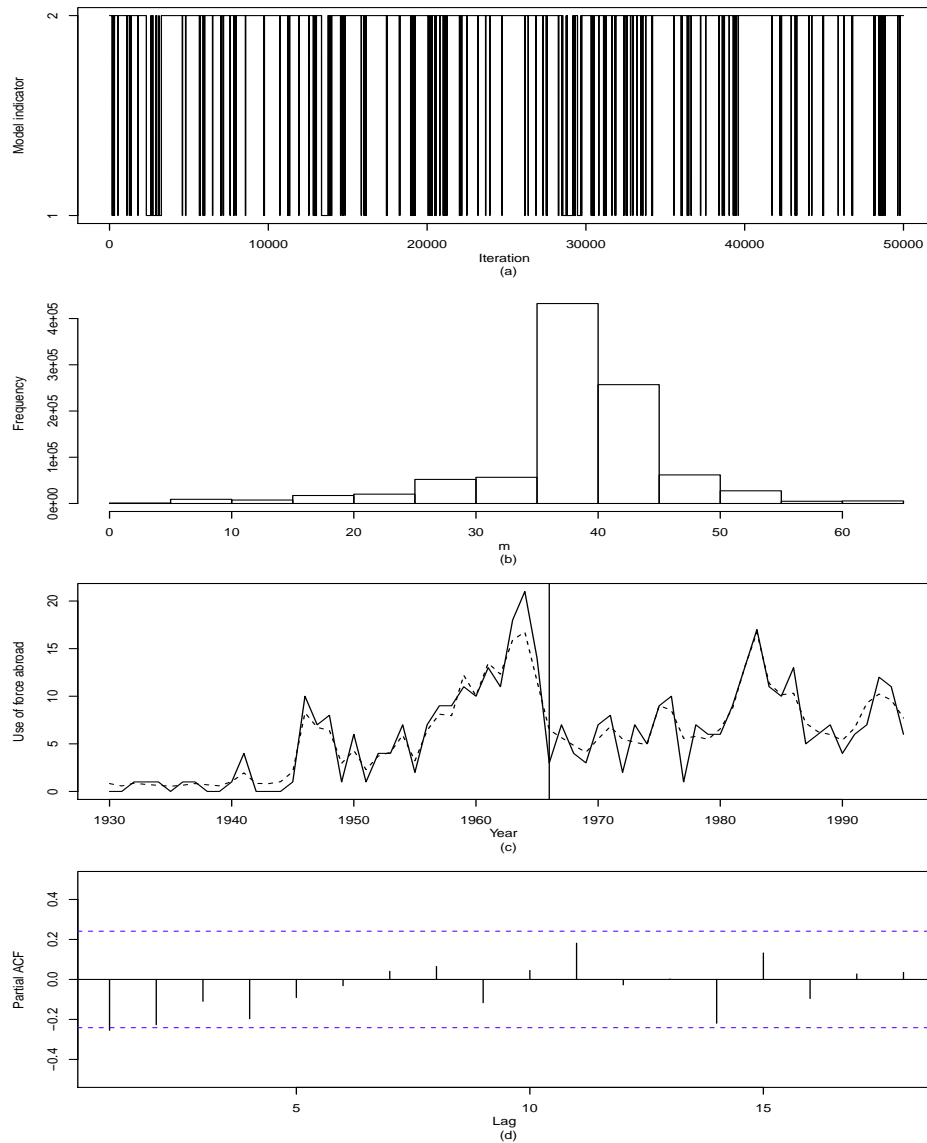| Parameter | Mean (s.e.) | Parameter | Mean (s.e.) |
|---|---|---|---|
| $\alpha_1$ | 0.8238 (0.2154) | $\alpha_2$ | 0.5995 (0.2931) |
| $\tau_1$ | 4.4472 (1.7884) | $\tau_2$ | 7.3936 (2.4154) |
| Intercept $(\beta_{0,1})$ | 0.6995 (0.2198) | Intercept $(\beta_{0,2})$ | 1.1466 (0.2762) |
| $t/n$ $(\beta_{1,1})$ | 0.5805 (0.3472) | $t/n$ $(\beta_{1,2})$ | 0.7985 (0.2596) |
| UNEMP $(\beta_{3,1})$ | 0.0627 (0.1153) | UNEMP $(\beta_{3,2})$ | 0.7529 (0.1449) |

Figure 5: Presidential uses of force data. (a) Trace plot of the model indicator for the last 50,000 iterations: 1 = no-changepoint model and 2 = changepoint model. (b) Histogram of the posterior distribution of $m$. (c) Observed data (solid line) vs posterior means of the fitted values (dashed line). The vertical line indicates the mode of the changepoint. (d) Partial autocorrelation plot of the residuals. The dotted lines indicate the $\pm 2$ standard error confidence limits.