Garth Holloway (UK)

On the environmental valuation of catastrophe: revisiting mining disasters in the United Kingdom, 1851-1962

Abstract

In the context of environmental valuation of natural disasters, an important component of the evaluation procedure lies in determining the periodicity of events. This paper explores alternative methodologies for determining such periodicity, illustrating the advantages and the disadvantages of the separate methods and their comparative predictions. The procedures employ Bayesian inference and explore recent advances in computational aspects of mixtures methodology. The procedures are applied to the classic data set of Maguire et al. (Biometrika, 1952) which was subsequently updated by Jarrett (Biometrika, 1979) and which comprise the seminal investigations examining the periodicity of mining disasters within the United Kingdom, 1851-1962.

Keywords: environmental valuation, catastrophe, mining disasters, United Kingdom, mixtures methodology, robust Bayesian inference.
JEL Classifications: N53, N54, Q54, Q55, C11, C18.

Introduction

One important issue overarching many facets of empirical enquiry is the notion of parameter stability – the notion that the quantities being estimated are constant throughout subsets of the population from which the empirical sample is drawn. Such stability and departures from it are important in the context of computations of welfare loss. This paper considers one example of welfare loss where agents are subject to changes in the parameter profile governing data generation. Specifically, the problem at hand concerns assessing and comparing welfare losses when the number of disasters, and their duration and frequency across the sample space change. Although this aspect of the calculations is only one of perhaps many such aspects of loss calculations, it is one important dimension. The topic is under-researched despite advances in recent decades in applied statistical procedures that are capable of assessing such changes. In the specific context of Bayesian empirical inference, the focus of this paper, a number of techniques have been devised that permit the parameters generating subsets of the sample data to vary. These methodologies include the so-called mixture-modeling framework, developed independently by Lavine and West (1992), Diebolt and Robert (1994) among others; and the hierarchical methodology showcased in the classic paper by Lindley and Smith (1972). The methodology that is commonly used in situations in which parameters undergo structural change is voluminous. A comprehensive review of that literature lies beyond the scope of the present effort. However, the multiple-change point model estimator developed by Chib (1998), which has remained hitherto unexploited in most agricultural-economic and environmental-economic settings is a natural candidate against which to consider alternative methodological formulations. As usual, different methods bring with them disadvantages as well as advantages and it is useful, instructive, and illuminating to study these constructions against one another (see Appendix).

While our main contribution is to revisit aspects of the multiple-change point methodology that we think should have wider exposure within the environmental-economic and agricultural-economic literatures, we are also concerned with problematic aspects of the methodlogy. And we wish to illuminate them.

In the context of disaster management and catastrophe calculations, we are fortunate to have available a detailed set of disaster records for mining accidents in the United Kingdom, 1851-1962. Not only are the data available (Maguire et al., 1952), updated (Jarrett, 1979) and are therefore amenable to investigation; they have been widely examined in previous work (Raftery and Akman, 1986; Worsley, 1986; Siegmund, 1988; Carlin et al., 1992; and Chib, 1998). This makes new insights on the same data especially illuminating.

In the next section of the paper we outline the basic ideas related to catastrophe calculation within a 'moving-target' setting. In section 2 we re-examine details pertaining to the mining disaster data. In section 3 we introduce notation and the basic densities that we use throughout the paper. In section 4 we present and discuss alternative methodology. In section 5 we present robust inference relating to the number of data-generating regimes and in section 6 we present empirical results. The paper concludes with some suggestion for future research.

1. Simplified catastrophe calculations

The problem at hand (Neumayer et. al, 2013) , when considered in its simplified canonical form, pertains to the computations of welfare losses, for which we employ the symbol 'λ', in the face of catastrophic disaster, for which we use the symbol, 'δ'. In its simplest setting, we seek to compute the usual compensating valuation, which is available from comparing the amount of monetary income, say, 'm' that individuals are willing to relinquish, in return for forestalling the disaster. In the usual symbolic terms, then, the amount we seek is the quantity, 'ω' that equates the individual's utility in the absence of disaster with the utility during disaster, with quantity ω acting as the so-called 'compensation'. In other words, we seek to measure the quantity ω that solves

$$V(m, \lambda) = V(m + \omega, \lambda + \delta), \tag{1}$$

In this representation, it is clear that in order to compute ω we must have available the (indirect) utility functions of the (representative) individual, $V(\cdot)$; the amount of money they consume as income, m, the level of disaster or 'loss', λ, that is present within the economy; and the amount by which it is changed by, or, 'δ,' during a disaster. In practice, the investigator has available neither, $V(\cdot)$, which may vary throughout the population of 'N' individuals, say, $V_i(\cdot)$, $i = 1, 2, .., N$; leading us to consider, aggregates, $\omega = \Sigma_i\ \omega_i$, which may vary in a complicated fashion, depending upon, the distribution of m_i, $i = 1, 2, .., N$; the distribution of $V_i(\cdot)$, $i = 1, 2, .., N$; as well as the total number of individuals, N, who comprise the population. Despite these complications, it is not unreasonable to consider a simplified relationship, between the all-important quantitites, ω, λ, δ, which are implicitly defined by the N conditions

$$\Sigma_i\ V_i(m_i, \lambda) = \Sigma_i\ V_i(m_i + \omega_i, \lambda + \delta), \tag{2}$$

The conditions (1) and (2) make very clear that, whether we consider the welfare loss of a representative individual, as we would do in the simplest case; or whether we seek to compute the distribution of the desired quantities throughout the entire population; both quantities, ω, on the one hand, and $\Sigma\omega_i$ on the other hand, will depend upon the important quantities, yet unknown, λ and δ.

Our analysis, therefore, and in what follows, will, of course depend crucially upon the magnitude of these two unknown quantities. We believe, that it is not unreasonable to consider a functional relationship, that is defined implicitly by the relations (1) and (2), and that the amount of money individuals would be willing to pay in order to forestall disaster, is monotonically increasing in the level of disaster, δ. In this setting, we may write, therefore

$$\omega = f(\delta, \lambda, m_1, m_2, .., m_N). \tag{3}$$

and assume that the income amount, ω, is a monotonically increasing function of the level of disaster, δ. So, we write, $f_\delta \geq 0$, in order to denote this relationship.

In order to further emphasize the importance of the loss amount, ω, the disaster level, δ, and the calculations that the social accountant should wish to make; consider their depiction, graphically, as presented in Figure 1.

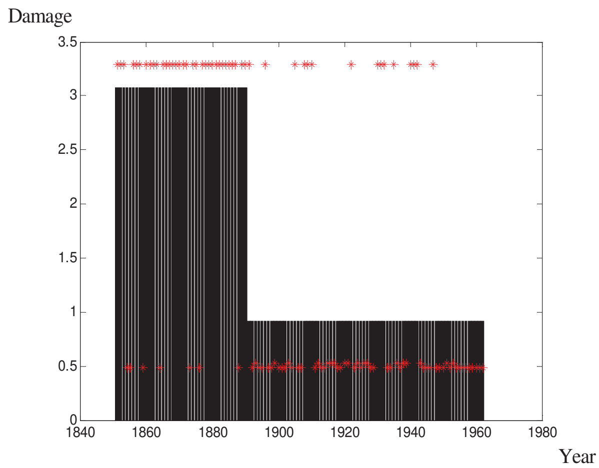


Fig. 1. Simple conceptual framework for evaluating damages over time periods

In the figure, the amount of loss is depicted vertically, and the amount of disaster at any given time, is also depicted vertically. On the horizontal axis, we present the time frame over which the calculations must be made. This depiction, therefore, has, in essence, two important dimensions to the calculation, namely the vertical level and the horizontal level, with each pertaining to important components of the welfare calculation. Within the figure, we consider two distinct regimes that may have arisen, in reality, due to the nature of extreme events, due to prolonged periods of minor disaster, and due to periods of tranquility. We make clear, however, that due to the two dimensions of the problem, their combinations across time and across loss horizons, and their variability across time and event, alternative loss calculations will, inevitably arise.

The figure depicts two distinct regimes, which we itemize by shaded rectangles, and it also contains reference to other points in the two-dimensional space. We will return to these points, highlighted by red asterisks, subsequently. Presently we simply summarize the main points of this section, by stating three main conclusions. First, given preferences and income, it is conceptually very straight-forward to calculate loss, if only we had the individual preference relations of the individuals concerned. Second, in the absence of such detail, we can conceive of an amount of loss, or a willingness to pay to forestall loss, that increases with the magnitude of a disaster level. And, third, most importantly, given the actual magnitude of loss within any given period, the duration and timing of the loss situation, can have very important ramifications for welfare loss calculations in the presence of disaster.

For these reasons, it is highly desirable to have a robust facility for the calculation of such amounts, which depend crucially on the location and scales of key parameters, but also upon the period of duration of change in multiple regime settings just as the one in which graphics in Figure 1, portray. In this context, a key objective of this paper is to present such a formal methodology for estimating the duration and the periodicity of disaster. And we explore the value of robust methodology for computing regime change in the face of disaster, against the backdrop of a highly-visited context, for which the data are available, namely mining disasters within the United Kingdom, 1851-1962.

2. Revisiting mining disasters in the United Kingdom, 1851-1962

Table 1 (see Appendix) presents the mining disaster data which we use to explore alternatives to conventional methodology estimating multiple regime changes across the sample space. By way of background, these data were originally collected by Maguire et al (1952), extended and corrected by Jarett et al. (1979); have been analyzed by frequentist methods by Wosley (1986) and by Siegmud (1988); and using the Bayesian approach, by Raftery and Akman (1986); then Carlin, Gelfand and Smith (1992); and, more recently, by Chib (1995). We report the data series in alternative graphical forms in Figure 2, Figure 3, Figure 4 and in Figure 5 (see Appendix) and we do this for three reasons. First, we wish to present to the reader a fairly comprehensive view of the number of disasters over the period, both marginally and cumulatively, and some basic idea about the way in which the frequency of disasters has evolved over time. Second, some inference, although casual and empirical, is available from careful examinations of the graphical outputs in the figures and this forms the mainstay of more detailed, formal enquiry. Third, we wish to illustrate a very important point with respect to the sequencing of the data which affects inference greatly and forms a natural evolution for the presentation of our investigation of the statistical basis underlying the number of mining disasters within the United Kingdom, 1851-1962. In order to investigate them, some notation will prove useful.

3. Notation

By way of notation we use lower-case Greek and Roman numerals to reference scalar quantities, use emboldened lower-case symbols to reference vectors and use emboldened upper-case symbols to reference matrix quantities. Thus, let $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, .., \theta_N)'$ denote a vector of parameters of interest, where ‘′’ denotes the ‘transpose’ of the column vector $\boldsymbol{\theta}$; $\pi(\boldsymbol{\theta})$ denotes the prior probability density function (*pdf*) for $\boldsymbol{\theta}$; and $\pi(\boldsymbol{\theta}|y)$ the posterior pdf for $\boldsymbol{\theta}$; where $y \equiv (y_1, y_2, .., y_N)'$ denotes data. Frequently, we reference the data generating model $f(y|\boldsymbol{\theta})$, which is the likelihood function when viewed as a function of $\boldsymbol{\theta}$ and, sometimes, make use of variants of the $f(\cdot|\cdot)$ notation in order to reference particular probability density functions. Occasionally we find it useful to reference just the variable part of the density (integrating constant excluded) in which case we use the symbol ‘$\propto$’ to denote ‘is proportional to.’ In view of the prior-to-posterior conjugacy shared by each model that we consider, we adopt the notational convention employed by Drèze and Richard (1983) wherein postscripts indicated ‘^{o}’ reflect prior information and postscripts indicated ‘*’ reflect posterior information; accordingly $f(\theta|\theta^o) \equiv \pi(\theta)$ and $f(\theta|\theta^*) \equiv \pi(\theta|y)$. Using, generically, ‘$x$,’ as their argument, the several pdfs that we apply are, respectively, the Dirichlet distribution $f^D(x|\wp) \equiv \Gamma(\Sigma_i\ \alpha_i) \times \Pi_i\ \Gamma(\alpha_i)^{-1} \times \Pi_i \wp_i^{\alpha i-1}$, $x \in [0, 1]$, where $\Gamma(\cdot)$ denotes the gamma function as described for example in Mood et al. (1972, p. 534); the Multinomial distribution $f^M(x|\wp) \equiv N! \times \Pi_i\ (x_i!)^{-1} \times \Pi_i \wp_i^{xi}$, $x_i = 0, 1, 2, .., N$; the Poisson pdf, $f^P(x|\lambda) \equiv \exp(-\lambda) \times \lambda^x \times (x!)^{-1}$, where $x = 0, 1, 2, \ldots$; and, finally, the Gamma pdf $f^G(x|r, \lambda) \equiv \lambda \times \Gamma(r) \times (\lambda x)^{r-1} \times \exp(-\lambda x)$, where x satisfies $x \in (0,+\infty]$.

4. Methods

Our major focus is the change-point methodology developed previously by Chib (1995), which is a very powerful unconstrained method for assessing the extent of regime change within a time series such as the one we are interested in considering here and which is portrayed, in alternative forms, within the figures (Figure 2-Figure 5). It’s basis, however, is the well-known Markov-switching model, which has been used extensively to study persistence in economic time series. However the starting point for both models is the basic mixtures model which has been popularized substantially with the advent of the Gibbs sampler (Gelfand and Smith, 1990; Gelfand, 2000). Before introducing, briefly, these three alternative methodologies, it is useful to investigate the properties of the basic data generating entity which provides the fulcrum in empirical investigation. This is the Poisson density.

4.1. Poisson analysis of the mining disaster data. The Poisson probability density function is a single-parameter density function for the counts in which the expected number of mining disasters but also their variance are equal to the value of the parameter 'λ'. In this context, we sometimes ask, in Bayesian investigations, whether there exists a conjugate situation, in which both the prior pdf and the posterior pdf have the same functional form. It transpires that the conjugate distribution for the Poisson data generating model is the Gamma distribution. Simply put, we can view the posterior distribution of the Poisson sampling or data-generating process as emanating from the Gamma prior distribution for the number of coal-mine disasters; combine it with the Poisson sampling process; and generate a Gamma posterior. Thus the Gamma prior and the Gamma posterior do indeed satisfy this property of conjugacy. In particular, if we allow ourselves to consider that the random sampling process evolves according to the steps $\lambda \sim f^G(\lambda|\alpha_o,\beta_o)$, $y \sim f^P(y|\lambda)$, we arrive at $\lambda \sim f^G(\lambda|y,\alpha_*,\beta_*)$, when, combined through Bayes rule. Thus, the two forms $\lambda \sim f^G(\lambda|\alpha_o,\beta_o)$, on the one hand and $\lambda \sim f^G(\lambda|\mathbf{y}, \alpha_*,\beta_*)$, on the other hand are indeed from the same conjugate family, and it remains to specify the precise relationship between the parameters α_o and β_o within the prior process $\lambda \sim f^G(\lambda|\alpha_o,\beta_o)$; and α_* and β_* within the posterior process $\lambda \sim f^G(\lambda|\mathbf{y}, \alpha_*,\beta_*)$. It occurs that the respective relations

$$\alpha_* \equiv \alpha_o + \Sigma_i\, y_i, \tag{4}$$

and

$$\beta_* \equiv \beta_o + N, \tag{5}$$

evolve. And when applied to the mining disaster data in table 1, we find that the posterior distribution, depicted in Figure 4 evolves. In order to summarize matters, because the posterior distribution for λ is in the Gamma form, which has a well defined mean, $E\{\lambda\} = \alpha_*/\beta_*$, and a well-defined variance $\mathrm{Var}\{\lambda\} = \alpha_*/\beta_*^2$, we know immediately, the form of the posterior distribution for λ, once we specify parameters for the prior pdf. The prior pdf parameters we employ are $\alpha_o \equiv 1$ and $\beta_o \equiv 1$. We use the principal, throughout this paper to apply weak but proper priors, and, so, consider, then to maintain the values $\alpha_o \equiv 1$ and $\beta_o \equiv 1$, throughout the entire analysis. In this way results are comparable and unaffected by changing the prior. Using these prior density values in the formulae in (4) and (5) we can compute that the posterior mean of the Poisson parameter is $E\{\lambda|y\} = 1.70$ and the posterior variance is $\mathrm{Var}\{\lambda|y\} = 0.02$. And these 'location' and 'scale' metrics are evident in Figure 4. We discuss Figure 5, subsequently.

The conjugate Gamma-Poisson-Gamma sampling model is used throughout the remainder of the paper, and we now ask, specifically, the nature of the distribution of the parameter and whether any change has occurred over the sample period. In order to examine this phenomena, we will analyze the data again using the Chib (1998) change-point methodology, which is essentially a special case of the Markov-switching methodology and is, in turn, a special case of the more general, though simpler, finite mixtures methodology. For this reason, we consider, in turn, the basic mixtures methodology and its estimating algorithm; the Markov-switching methodology; and, finally, the multiple change-point methodology.

4.2. Finite-mixtures methodology. The basic and fundamental finite mixtures methodology evolves from considering that the data $y_1, y_2, .., y_N$ are a sequence of random generations from one component, say component 'k,' of a mixture of Poisson densities where we assume that each density arises naturally as it does in the single-regime case, just covered, but that the data are drawn with unknown probabilities $\omega_1, \omega_2, .., \omega_K$, in the sample space. If we assume, also that each probability is itself the product of a draw from a Dirichlet prior distribution, and then augment the observed data likelihood with unknown and unobservable classification data, a convenient formulation arises, wherein, following a fairly well-trodden path (Tanner and Wong (1987), Lavine and West (1992), Diebolt and Robert (1994), Chib (1995), among others) we are then able to exploit convenient mathematical properties of the model which lend themselves to a very straight-forward Gibs sampling algorithm consisting of the steps:

Conditional on classification and component specific Poisson parameters, say, $\lambda_1, \lambda_2, .., \lambda_K$, when K regimes exits, draw the mixing weights from a Dirichlet distribution; conditional upon the mixing weights and the value of the Poisson parameters draw the classification data from a Multinomial distribution; and, third, conditional on the classification data and the mixing weights, draw the parameters $\lambda_1, \lambda_2, .., \lambda_K$, from their component specific data.

Under normal circumstances the algorithm is very robust and leads to robust estimates of the unknowns and allows one to draw inferences about the unknown parameters governing the independent regimes. Numerous applications exist within the literature and reviewing it lies beyond the scope of our investigation. One very useful and instructive presentation is Dellaportas (1998).

4.3. Markov-switching methodology. The Markov switching methodology is a very small modification of the basic finite mixtures algorithm, in which, unlike that basic algorithm, the data are not treated as fully independent, or 'exchangeable' across the sample space. The algorithm essentially differs in the draw for the mixing weights which now follow a Markov process. However, in conjugate settings, such as the present one, the steps are essentially the same. In particular we have the following steps evolving:

Conditional on classification and component specific Poisson parameters, say, λ_1, λ_2, .., λ_K, when K regimes exits, draw the mixing weights from a Markov distribution which, for identification purposes is time-homogeneous, irreducible and aperiodic; conditional upon the mixing weights and the value of the Poisson parameters draw the classification data from a Multinomial distribution; and, third, conditional on the classification data and the mixing weights, draw the parameters λ_1, λ_2, .., λ_K, from their component specific data.

An example of the algorithm is presented in the work of Chib (1996).

4.4. Multiple-change-point methodology. The change-point model is a special case of the Markov switching model in which the Markov regime for the mixing weights is forced to move in a single direction throughout the sample, for $t = 1$, $t = 2$, $t = 3$, … and so on. It consists of an algorithm that is almost identical to the Markov model, with the inclusion of this basic restriction on the switching across regimes. Consequently, the change-point methodology is now easily developed based on its roots in the finite mixtures methodology and in the Markov-switching setting.

Let 'm' denote the number of change-points permitted by the investigator. Following Chib (1998), let $Y_n \equiv \{y_1, y_2, .., y_n\}$ denote a time series such that the density of y_t given Y_{t-1} depends on some parameter ξ_t whose value changes at unspecified points in time $T \equiv \{\tau_1, \tau_2, .., \tau_m\}$ and remains constant otherwise where $\tau_1 > 1$ and $\tau_m < n$. With 'm' so-called 'breakpoints', the setup is:

$$\xi_t = \begin{cases} \theta_1 & if \quad t \le \tau_1, \\ \theta_2 & if \quad \tau_1 < t \le \tau_2, \\ \ldots & \ldots \qquad \ldots \\ \theta_m & if \quad \tau_{m-1} < t \le \tau_m, \\ \theta_{m+1} & if \quad \tau_m < t \le n. \end{cases} \tag{6}$$

In terms of the basic mixtures and Markov-switching algorithms, there are two additional quantities that need to be estimated, namely the probability of remaining in the present regime, or $\wp \equiv \{\wp_1, \wp_2, .., \wp_{m-1}, \wp_m\}$, and the sequence of states $S \equiv \{s_1, s_2, .., s_n\}$ throughout the time series.

One fundamental contribution of Chib (1998) is to show that these two quantities are easily simulated by sampling from two additional steps setting $m = 1, 2, \ldots n$, respectively. Step 1: Draw $\wp_j$, $j = 1, 2, .., m$, from $\pi(\wp_j|) \propto f^{\beta}(\wp_j|a + n_{ii}, b+1)$, where parameters 'a' and 'b' evolve from the natural conjugate prior pdf $\pi(\wp_j) \equiv f^{\beta}(\wp_j|a, b)$ and n_{ii} denotes the total number of one-step transitions from state 'i' to state 'i' within the sample. Step 2: Draw s_t, $t = 1, 2, .., n$, from $\pi(s_t|\cdot) \propto f^b(s_t|p)$ with probability $p(s_t = k|Y_t, \Theta, \wp) \propto p(s_t = k|Y_{t-1}, \Theta, \wp) \times f(y_t|Y_{t-1}, \theta_k)$, where $f(\cdot)$ here denotes the sampling distribution for the data,.

Chib's (1998) multiple change-point methodology is attractive for at least four reasons. First, the 'new' change-point methodology overcomes the problem of dimensional intractability arising from the fact that the state vector, $S \equiv (s_1, s_2, .., s_T)'$ can have as many as M^T paths, where T denotes the number of observations in the time series and M denotes the number of alternative states. Second, unlike previous approaches, the 'new' methodology does not require the investigator to constrain the probabilities of change to be constant at each point in time (Chernoff and Zacks, 1964) or that the joint distribution of the parameters is exchangeable and is independent of the change points (Yao, 1984). Third, the new methodology is completely general, deriving posterior distributions for the state variables and the probabilities of a change at any particular point in time that are functionally independent of the sampling distribution. The latter feature bestows upon the new methodology a considerable degree of robustness and makes possible extensions of the basic framework that enables the spirit of enquiry to be targeted to a highly varied set of sampling structures.

5. Robust inference about the number of data generating regimes

One key feature shared by each of the three settings considered thus far (and one that is also relevant to the basic Poisson setting) is the assumption concerning the number of mixture components, or the integer value, 'M'. Indeed careful re-examination of the algorithms above sees that they are in fact each dependent on the value M. In realistic sample settings, however, the investigator will usually have incomplete information about the number of regimes and it is desirous, therefore, to estimate M. Several alternatives exist, but one which is potentially desirous, namely Chib (1995) is problematic, due to the fact that the 'labels' within the mixing regimes are not identifiable.

This situation has received some considerable attention in the literature of late, and is covered rather comprehensively in Berkhof, van Mechelen and Gelman (2003) and the literature cited therein. The basic problem arises as follows. Within any mixtures setting, there arise $M!$ modes which must be evaluated during estimation. Not only during estimation is this problematic, but also when one wishes to evaluate the evidence in favor of a particular model and therefore one needs to estimate the all-important marginal likelihood, as in Chib (1995).

The latter aspect of estimation is particularly troublesome when it is necessary to determine the number of regimes to which the finite mixtures, Markov-switching or multiple-change-point methods are applied. We conjecture that this is almost always the case. And it is unsurprising that attempts have been made to develop methodology, alternative methodology, for this specific purpose.

Initial work on estimating the number of mixing components evolves from the notion of 'reversible-jump' MCMC (Markov Chain Monte Carlo) methods as availed by Green (1995). This approach has proven quite robust in a variety of settings outside of the mixtures genre. However a key contribution in the evolution of technique is the seminal paper by Green and Richardson (1997) where they show, specifically how the number of mixtures can be determined in principal, and in applications. The Green and Richardson (1997) approach has now become quite standard in recent years and is usually referred to under the general umbrella of 'adaptive MCMC.' The term 'adaptive' arises because the dimensional space of the unknown quantities is adjusted – both reduced and expanded – endogenously, throughout the sampling chain. This aspect of the estimation is particularly noteworthy.

Complications in storage, in post sample processing and in collating the output of the MCMC sample arise and, generally speaking, lie above and beyond the demands in the typical finite mixtures methodology (Lavine and West, 1992; Diebolt and Robert, 1994; Chib, 1995; Dellaportas, 1998). Within these applications the number of components to the mixture are fixed. It serves to emphasize that these computations are performed under the conditional assumption of a fixed number of mixture components (Green and Richardson, 2002).

Modifications that require the so-called 'detailed-balance' condition to prevail are especially complex in the Richardson and Green (1997) approach. 'Detailed balance' is the condition that enables the MCMC chain to regress and navigate within the parameter space but also within the space of dimensions of the unknowns defining the parameter space. One crucial issue arises within the context of mixtures methodology in the adaptive sampling genre. The issue is how to expand and contract the mixing weight parameters when the chain wishes to contract or expand across the components space. An alternative methodology, which is built directly upon extensions of the Richardson and Green (1997) methodology, handles this issue in a particularly appealing manner and appears to suffer less complexity due to a refinement of the likelihood used in estimation. The approach is presented in Stephens (2000), about which, it is noted:

Richardson and Green present a method of performing Bayesian analysis of data from a finite mixture distribution with an unknown number of components. Their method is a Markov Chain Monte Carlo (MCMC) approach, which makes use of the "reversible jump" methodology described by Green. We describe an alternative MCMC method which views the parameters of the model as a marked (point) process, extending methods suggested by Ripley to create a Markov birth-death process with an appropriate stationary distribution. Our method is easy to implement, even in the case of data in more than one dimension, and we illustrate it on both univariate and bivariate data. There appears to be considerable potential for applying these ideas to other contexts as an alternative to more general reversible-jump methods and we conclude with a brief discussion of how this might be achieved (Stephens, 2000, p. 40).

It is particularly noteworthy that, in closing, Stephens (2000, p. 67) lays out the extension of the finite mixtures settings in order to model a multiple change-point situation. However, to our knowledge, such an algorithm has yet to be applied in any empirical setting.

Our endeavor is eventually to apply the full change-point methodology on the coal mine disaster data, for which a first step is to consider the closely related issue of the number of components in the basic finite mixtures setting. In what follows, we investigate, through application of the Stephens (2000) approach, the number of components existing in the coal-mine disasters dataset collected for the United Kingdom, 1851-1962.

6. Empirical models and results

The empirical models implemented are a simplified version of the model applied by Chib (1995) to estimate finite mixtures in the aurora borealis, and Chib (1998), who applies the standard change-point methodology on the coal-mine data we employ. The reader is referred to Chib (1995) and Chib (1998) for computational details, but the algorithms proceed, essentially, as outlined, at least superficially, above, in the methods section. Draws, only from, respectively, Multinomial, Gamma and Dirichlet distributions are all that are required, and it is quite straight-forward to initiate simple Gibbs sampling algorithms for our purpose. Additional details are available from the author upon request as are the computer codes used to produce all of the results we are about to digest.

One issue arising in the context of change-point investigations, which is important, is *the sequence of observations within the sample*. In time-series applications there is of course a very natural sequence. This situation does not arise in settings in which the observations are independent and arise as random draws throughout the time series, that is, draws from a pre-specified number of mixtures components. This point is particularly noteworthy when the issue of marginal likelihood computation as in Chib (1995) arises. In this context, the way in which the algorithm is extended can lead to problematic findings as the following analysis illustrates.

Recalling that Figures 2 through 5 (see Appendix) present alternative descriptions of the coal mine disaster data. In Figure 2 we presented the coal-mine disasters as a bar chart and in Figure 3 we presented the cumulative distribution of the disasters over time. From Figure 2 we observe a highly varied pattern of scatter of disasters during the period 1851-until about 1892 and then a decline and a rise again and then eventual decline. This pattern of sequencing throughout the series is evident from the cumulative perspective in Figure 3 where we encounter a fairly steep rise in accumulation during the 1851-1892 period followed by a lessening in steepness of rise and then by a flattening in the accumulations. It is evident at least from this casual empirical evidence that there may be changes in the underlying components of the coal-mine series and that change-point investigation is required.

A start point for such analysis is to identify what the data suggest if they are combined and estimated as a single regime. This is undertaken as above, where we detail that the conjugate Gamma posterior for the all-important Poisson parameter, yields an expected value and variance estimate corresponding to the location and sale report as evidence by Figure 4. The posterior distribution is in fact centered at the mean which is 1.70 and the variance is around 0.02.

We note that this conclusion is the same one that an investigator would arrive at using the entire sample and processing the data sequentially. And we note that this outcome is *independent of the order in which the data are processed*. Therefore, consider the alternative orderings of the disaster data as in Table 2 and in Figure 5. In Table 2 we present three such orderings, the first comprising the natural ordering of the time series; the second comprising the time series in reverse order; and the third comprising the time series in a random order; and consider inference about the all-important Poisson parameter 'λ' when these transitions are processed in these three, respective orderings. The answer is portrayed in the three distinct series plotted in Figure 5. In the first case, the natural case, in which the time series are processed in natural order, we arrive at the series plotted in red. There is evidence of an initial oscillation; followed by a period of increase; stability; and then a gradual decline in the estimates; culminating in the eventual point estimate corresponding with the midpoint of the density projected in Figure 4. We note, in passing, that methodology exists for making gradual switching regression and is available and has been used to model gradual structural changes in meat demand (Moschini and Meilke, 1997), in the classical context; and in production (Tsurumi et al., 1986), in Bayesian settings. We focus attentions, however, on the standard change point and mixtures methodologies. Thus, from sequentially updating the Poisson parameter, we conclude that there has been a gradual decline in disasters from about the first 40 observations within the sample.

In contrast, focusing upon the blue line in Figure 5, we observe initial oscillation with a gradual increase in trend; rising steeply towards the end of the series until, at the last observation, expectations coincide with the midpoint of the posterior mean reported in Figure 4.

And, when the data are ordered randomly, we observe the trend much akin to one we expect to observe in a single regime setting. Thus, the question arises about the number of regimes prevailing in the data generating environment and the manner in which they are processed. In the first and second settings there is clear evidence of regime switching, but in the third there is not. Thus, how many regimes evolve from formal analysis?

Figure 6 provides an answer to this question using the Stephens (2000) methodology. It is worth mentioning that in order to implement the Stephens method one important user-supplied piece of information must be incorporated into the estimation process. This piece of information is the so-called 'birth rate' at which new mixtures components are created. The figure depicts the posterior probabilities assigned to alternative sets of mixture components with the darker shade denoting a birth rate of '1' the midrange shade depicting a birth rate of '2' and the lighter shade depicting a birth rate of '3.' The probabilities suggest a fairly even split between probabilities placed on the mixture components '2' and '3' with negligible mass deposited elsewhere. And we conclude, therefore, that there is fairly robust evidence to support the conclusion that the number of components underlying the mining disaster data is around *two* or *three*.

Another issue arising, which follows logically and naturally from the first, is how to classify the observations into the separate regimes. This concern is especially interesting due to the results of previous work where Chib (1998) and others seems to find fairly strong support that a significant structural change occurred within the data series at or around time period 41, which coincides with the year 1892. They find two regimes governed by Poisson parameter posterior mean (standard deviation) values of $\lambda^{(1)} = 3.119$ (0.286) and $\lambda^{(2)} = 0.957$ (0.120), respectively. In order to investigate this finding in more detail, we implement a 'label-free' procedure, in which the classifications are permitted to roam throughout the sample and we classify the various observations as coinciding with one or other of the two-component, or three-component regimes, running separate MCMC chains in each case. The procedure with which we produce classification is a subtle matter circumventing the problem – the over-arching dilemma confronting 'classification' – that 'labeling' which is prima facie necessary in a classification exercise generates most problematic findings. It is outline, briefly, in Holloway (2013) and lies beyond the scope of present interests.

Turning now to the classification of the time series and its importance for the over-arching loss calculations, the reader is returned to Figure 1. In the figure, the values of the loss calculations using the Chib (1998) methodology are compared and contrasted with those derived under the Stephens (2003) method. The former are depicted by the shaded, enclosed area and the latter by the scatter entries (in red) superimposed on the former. We make two observations in conclusion. First, a significant difference in regime selection components governing the coal-mine disaster data appears to exist. Second, it leads to considerable bias in welfare loss calculations in the face of disaster.

Concluding comments

This study highlights the usefulness, versatility and general dexterity of an extremely robust methodology for assessing structural change in time series but more generally, parameter stability in a wide variety of data generating settings. The methodology is the endogenous regime determination point process developed by Stephens (2000). It remains underutilized in applied environmental settings. Our results and examples in the simple illustrations presented here suggest that the procedure should have broader exposure. Further work with the change-point methodology presented within this paper is likely to yield robust inferences in a variety of setting that are of fundamental importance to statistical enquiries within environmental economics.

Acknowledgements

This paper is dedicated to one colleague whom, in ill-health, temporarily, we hope, is recovering from her own personal 'disaster.' The author is grateful to her for her support. He is also very grateful to his students for numerous conversations about the topic. And he is also grateful to AFIT (The Agriculture and Food Investigation Team in The School of Agriculture Policy and Development at The University of Reading) for 'enabling' the computer-intensive components of this research to be executed efficiently. All of the programming codes and data are available from the author upon request and he alone is responsible for any remaining errors.

Basic references

Basic references on change point analysis at a level of analytical depth conversant with, say, Zellner (1971), include Shiryayev (1963), Chernoff and Zacks (1964), Hinkley (1970), Broemeling (1972), Ferreira (1975), Smith (1975), Holbert and Broemeling (1977), Chin Choy and Broemeling (1980), Menzefricke (1981), Booth and Smith (1982), Diaz (1982), Hsu (1982), Zacks (1983), Wolfe and Schechtman (1984), Yao (1984), Moen et al. (1985), Raftery and Akman (1986), Siegmund, 1986), Smith and Cook (1986), Worseley (1986), Carlin et al. (1995) and Chib (1998). Basic reference on Bayesian model selection, in general, include Zellner (1971), Carlin and Chib (1995), Chib (1995), Draper (1995), Green (1995), Brooks et al. (2003) and Walker (2012).

References

1. Booth, N.B. and Smith, A.F.M. (1982) A Bayesian approach to retrospective identification of change points, *Journal of Econometrics*, 19, pp. 7-22.
2. Broemeling, L. (1972). Bayesian procedures for detecting a change in a sequence of random variables, *Metron*, 30, pp. 1-14.
3. Brooks, S.P., P. Giudici and G.O. Roberts (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions, *Journal of the Royal Statistical Society*, Series B, 65 (1), pp. 3-55.
4. Carlin, B. and S. Chib (1995). Bayesian model choice via Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society*, Series B Methodological, 57 (3), pp. 473-484.

5. Carlin, B.P., Gelfand, A.E. and A.F.M. Smith (1992). Hierarchical Bayesian Analysis of Change-Point Problems, *Applied Statistics*, 41 (2), pp. 389-405.
6. Chernoff H., Zacks S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time, *The Annals of Mathematical Statistics*, 35, pp, 999-1018.
7. Chib S (1995). Marginal likelihood from the Gibbs output, *Journal of the Royal Statistical Society*, 90, pp. 1313-1321.
8. Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics*, 75, pp. 79-98.
9. Chib S. (1998). Estimation and comparison of multiple change-point models, *Journal of Econometrics*, 86, pp. 221-241.
10. Chin Choy, J. and Broemeling, L. (1980). Some Bayesian inferences for a changing linear model, *Technometrics*, 22, pp. 71-78.
11. Diaz, J. (1982). Bayesian detection of a change of scale parameter in sequences of independent gamma random variables, *Journal of Econometrics*, 19, pp. 23-29.
12. Diebolt J., Robert C.P. (1994). Bayesian estimation of finite mixture distributions, *Journal of the Royal Statistical Society*, B 56, pp. 363-375.
13. Draper, D. (1995). Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society*, Series B, Methodological, 57 (1), pp. 45-97.
14. Drèze, J.H., Richard, J.-F. (1983). Bayesian Analysis of Simultaneous Equations Systems in Handbook of Econometrics, in: Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, edition 1, volume 1, chapter 9, pp. 517-598. Amsterdam: Elsevier.
15. Ferreira, P.E. (1975). A Bayesian analysis of a switching regression model: a known number of regimes, *Journal of the American Statistical Association*, 70, pp. 370-374.
16. Gelfand, A.E. (2000). Gibbs Sampling, *Journal of the American Statistical Association*, 95, pp. 1300-1304.
17. Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, pp.98-409.
18. Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82, pp.711-732.
19. Green, P.J. and S. Richardson (2002). Hidden Markov models and disease mapping, *Journal of the American Satistical Association*, 97, pp. 1-16.
20. Hinkley, D.V. (1970). Inference about the change-point in a sequence of random variables, *Biometrika*, 57, pp. 1-17.
21. Holbert, D. and Broemeling, L.D. (1977). Bayesian inference related to shifting sequences and two-phase regression, *Communications in Statistics*, *A*, 6, pp. 265-275.
22. Holloway, G. (2013). A Classification Paradox. Unpublished manuscript.
23. Hsu, D.A. (1982). A Bayesian robust detection of shift in the risk structure of stock market returns, *Journal of the American Statistical Association*, 77, pp. 29-39.
24. Jarrett, R.G. (1979). A note on the intervals between coal-mining disasters, *Biometrika*, 66, pp. 191-193.
25. Lavine M., West M.D. (1992). A Bayesian method for classification and discrimination, *Canadian Journal of Statistics*, 20, pp. 451-461.
26. Lindley D.V., Smith A.F.M. (1972). Bayes estimates for the linear model, *Journal of the Royal Statistical Society*, B Met, 34, pp. 1-41.
27. Yao Y. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches, *The Annals of Mathematical Statistics*, 12, pp. 1424-1447.
28. Maguire, B.A., Pearson, E.S. and Wynn, A.H.A. (1952). The time intervals between industrial accidents, *Biometrika*, 38, pp. 168-180.
29. Menzefricke, U. (1981). A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point, *Applied Statistics*, 30, pp. 141-146.
30. Moen, D.H.,, Salazar, D. and Broemeling, L.D. (1985). Structural changes in multivariate regression models, *Communications in Statistics*, *A*, 14, pp. 1757-1768.
31. Mood A.M., Graybill F.A., Boes D.C. (1974). Introduction to the Theory of Statistics. New York: McGraw-Hill, third edition.
32. Neumayer, E., Plümper, T. and F. Barthel (2013). The Political Economy of Natural Disaster Damage, *Global Environmental Change* (in press).
33. Raftery, A.E. and Akman, V.E. (1986). Bayesian analysis of a Poisson process with a change-point, *Biornetrika*, 73, pp. 85-89.
34. Richardson, S. and P. Green (1997). On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society*, Series B, 59 (4), pp. 731-92. (with discussion).
35. Ripley, B. (1987). Stochastic Simulation. New York: Wiley.
36. Shiryayev, A.N. (1963). On optimum methods in quickest detection problems, *Theory of Probability and Its Applications*, 8, pp. 22-46.
37. Siegmund, D. (1986). Boundary crossing probabilities and statistical applications, *Annals of Statistics*, 14, pp. 361-404.
38. Siegmund, D. (1988). Confidence sets in change point problems, *International Statistical Review Annals of Statistics*, 56, pp. 31-48.
39. Smith, A.F.M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables, *Biometrika*, 62, pp. 407-416.

40. Smith, A.F.M. and Cook, D.G. (1980). Straight lines with a change-point: a Bayesian analysis of some renal transplant data, *Applied Statistics*, 29, pp. 180-189.
41. Tanner, M. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation (with discussion), *Journal of the American Statistical Association*, 82, pp. 528-550.
42. Walker, S. (2012). A Gibbs Sampling Alternative to reversible Jump MCMC. Mimeograph, Institute of Mathematics and Statistics, University of Kent.
43. Wolfe, D.A. and Schechtman, E. (1984). Nonparametric statistical procedures for the change point problem, *Journal of Statistical Planning and Inference*, 9, pp. 389-396.
44. Worsley, K.J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables, *Biometrika*, 73, pp. 91-104.
45. Zacks, S. (1983). Survey of classical and Bayesian approaches to the change point problem: fixed sample and sequential procedures of testing and estimation. In Recent Advances in Statistics: Herman Chernoff Festschrift, pp. 245-269. New York: Academic Press.
46. Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley Classics Library Edition.

Appendix

Table 1. Coal mining disasters in the United Kingdom, 1851-1962

Year	Number of disasters	Year	Number of disasters
1851	4	1907	0
1852	5	1908	3
1853	4	1909	2
1854	1	1910	2
1855	0	1911	0
1856	4	1912	1
1857	3	1913	1
1858	4	1914	1
1859	0	1915	0
1860	6	1916	1
1861	3	1917	0
1862	3	1918	1
1863	4	1919	0
1864	0	1920	0
1865	2	1921	0
1866	6	1922	2
1867	3	1923	1
1868	3	1924	0
1869	5	1925	0
1870	4	1926	0
1871	5	1927	1
1872	3	1928	1
1873	1	1929	0
1874	4	1930	2
1875	4	1931	3
1876	1	1932	3
1877	5	1933	1
1878	5	1934	1
1879	3	1935	2
1880	4	1936	1
1881	2	1937	1
1882	5	1938	1
1883	2	1939	1
1884	2	1940	2
1885	3	1941	4
1886	4	1942	2
1887	2	1943	0
1888	1	1944	0
1889	3	1945	0
1890	2	1946	1
1891	2	1947	4
1892	1	1948	0

Table 1 (cont.). Coal mining disasters in the United Kingdom, 1851-1962

Year	Number of disasters	Year	Number of disasters
1893	1	1949	0
1894	1	1950	0
1895	1	1951	1
1896	3	1952	0
1897	0	1953	0
1898	0	1954	0
1899	1	1955	0
1900	0	1956	0
1901	1	1957	1
1902	1	1958	0
1903	0	1959	0
1904	0	1960	1
1905	3	1961	0
1906	1	1962	1

Note: The data are from Maguire et al. (1952) corrected by Jarrett (1979).

Table 2. Year sequences of coal mining disasters in the United Kingdom, 1851-1962

One	Two	Three
1851	1962	1902
1852	1961	1958
1853	1960	1961
1854	1959	1906
1855	1958	1911
1856	1957	1857
1857	1956	1938
1858	1955	1868
1859	1954	1945
1860	1953	1920
1861	1952	1957
1862	1951	1941
1863	1950	1908
1864	1949	1890
1865	1948	1947
1866	1947	1896
1867	1946	1950
1868	1945	1879
1869	1944	1901
1870	1943	1954
1871	1942	1858
1872	1941	1895
1873	1940	1853
1874	1939	1891
1875	1938	1878
1876	1937	1873
1877	1936	1910
1878	1935	1888
1879	1934	1877
1880	1933	1919
1881	1932	1871
1882	1931	1925
1883	1930	1948
1884	1929	1867
1885	1928	1897
1886	1927	1881
1887	1926	1887

Table 2 (cont.). Year sequences of coal mining disasters in the United Kingdom, 1851-1962

One	Two	Three
1888	1925	1952
1889	1924	1883
1890	1923	1886
1891	1922	1893
1892	1921	1955
1893	1920	1923
1894	1919	1942
1895	1918	1860
1896	1917	1870
1897	1916	1889
1898	1915	1864
1899	1914	1930
1900	1913	1935
1901	1912	1884
1902	1911	1912
1903	1910	1900
1904	1909	1928
1905	1908	1924
1906	1907	1936
1907	1906	1937
1908	1905	1921
1909	1904	1856
1910	1903	1851
1911	1902	1903
1912	1901	1940
1913	1900	1854
1914	1899	1882
1915	1898	1909
1916	1897	1885
1917	1896	1862
1918	1895	1880
1919	1894	1917
1920	1893	1875
1921	1892	1944
1922	1891	1918
1923	1890	1852
1924	1889	1929
1925	1888	1939
1926	1887	1859
1927	1886	1943
1928	1885	1932
1929	1884	1892
1930	1883	1899
1931	1882	1876
1932	1881	1915
1933	1880	1872
1934	1879	1949
1935	1878	1866
1936	1877	1913
1937	1876	1956
1938	1875	1962
1939	1874	1927
1940	1873	1865
1941	1872	1874

Table 2 (cont.). Year sequences of coal mining disasters in the United Kingdom, 1851-1962

One	Two	Three
1942	1871	1934
1943	1870	1946
1944	1869	1922
1945	1868	1916
1946	1867	1914
1947	1866	1863
1948	1865	1855
1949	1864	1861
1950	1863	1960
1951	1862	1931
1952	1861	1959
1953	1860	1904
1954	1859	1907
1955	1858	1926
1956	1857	1864
1957	1856	1905
1958	1855	1951
1959	1854	1953
1960	1853	1933
1961	1852	1898
1962	1851	1869

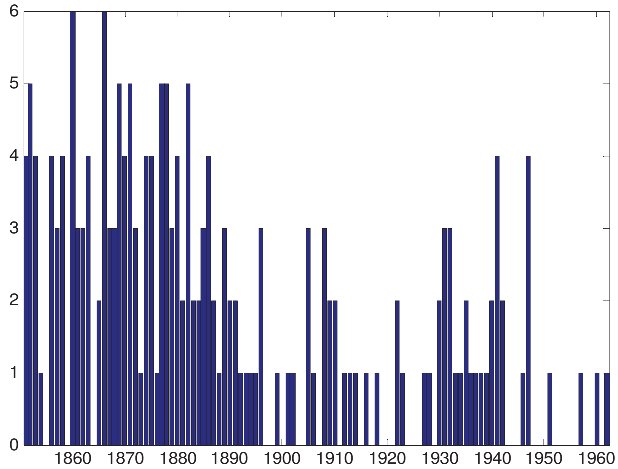


Fig. 2. Distribution of coal mining disasters in the United Kingdom, 1851-1962

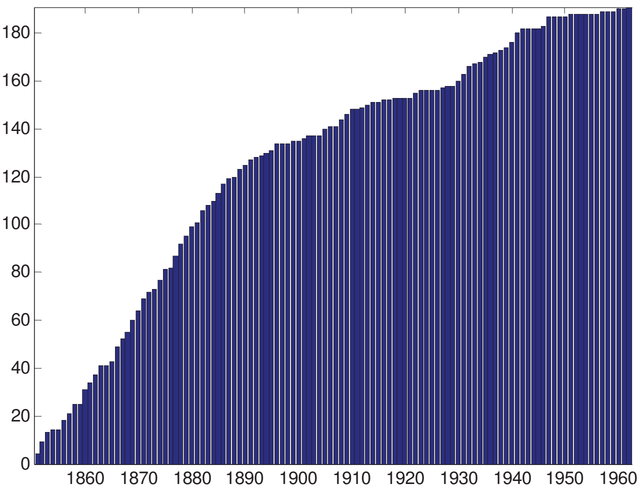
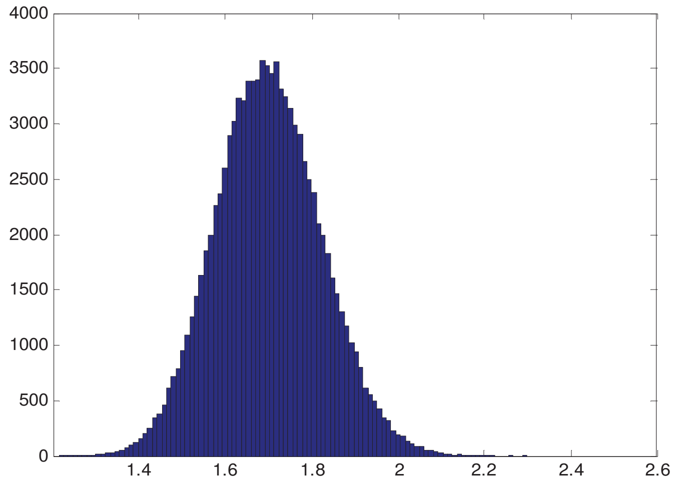
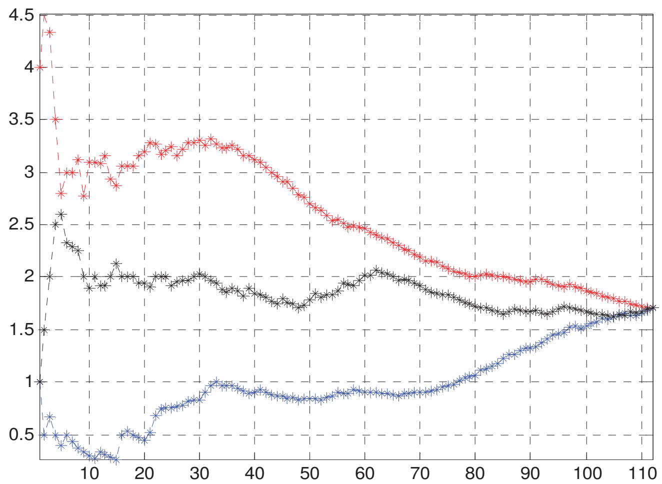


Fig. 3. Cumulative distribution of coal mining disasters in the United Kingdom, 1851-1962



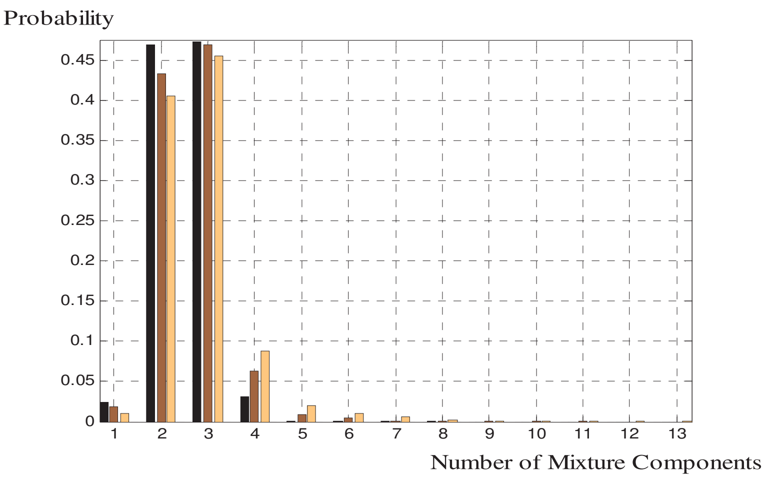
Note: The posterior mean is 1.70 and the posterior variance is 0.02.

Fig. 4. Posterior distribution of the Poisson parameter for the mining disaster data



Notes: The lighter shade line depicts the sequence moving through the data in the sequence 1851, 1852, …, 1961. The midrange shade line depicts the sequence moving in reverse order through the sample from years 1962, 1961, …, 1852, 1851. The darkest shade line depicts the sequence moving through the data in the order presented in the randomly selected order presented in Table 2.

Fig. 5. Sequential updating of the Poisson parameter through the sample of coal mine disasters in the United Kingdom, 1851-1962



Notes: The probabilities are reported for three separate birth rates, with the darkest bars depicting birth-rate = 1, the midrange shade depicting birth-rate = 2, and the lighter shade depicting birth-rate = 3.

Fig. 6. Posterior probabilities of the number of mixture components using the Stephens (2000) methodology