ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach

Michel Lubrano[a,*], Abdoul Aziz Junior Ndoye[b,1]

[a] Aix-Marseille University (AMSE) and GREQAM-CNRS & Ehess, Centre de La Vieille Charité, 2 rue de la Charité, 13002 Marseille, France

[b] Université d'Orléans, LEO (UMR 7322), Rue de Blois - BP 6739, 45067 Orléans Cedex 2, France

ARTICLE INFO

Article history:
Received 14 October 2013
Received in revised form 6 October 2014
Accepted 9 October 2014
Available online xxxx

Keywords:
Income distribution
Log-normal distribution
Inequality decomposition
Mixture models
Gibbs sampling
Label switching

ABSTRACT

The log-normal distribution is convenient for modelling the income distribution, and it offers an analytical expression for most inequality indices that depends only on the shape parameter of the associated Lorenz curve. A decomposable inequality index can be implemented in the framework of a finite mixture of log-normal distributions so that overall inequality can be decomposed into within-subgroup and between-subgroup components. Using a Bayesian approach and a Gibbs sampler, a Rao-Blackwellization can improve inference results on decomposable income inequality indices. The very nature of the economic question can provide prior information so as to distinguish between the income groups and construct an asymmetric prior density which can reduce label switching. Data from the UK Family Expenditure Survey (FES) (1979 to 1996) are used in an extended empirical application.

1. Introduction

Inequality within countries has increased a lot in recent years due to the effects of globalization and biased technological changes (see e.g. Bourguignon, 2012), leading to an elongation of the right-hand tails of the different national income distributions. If we want to make inference about these income distributions, it is more common to let the data speak for themselves using a non-parametric approach than to impose a restricted parametric model (Marron and Schmitz, 1992). However, a non-parametric approach can lead to unreliable inference in the presence of heavy tails as explained in theoretical terms in Bahadur and Savage (1956). There is thus an interest in considering new families of parametric densities that have been proposed with the aim of providing a greater flexibility in modelling of the tails, a point where non-parametric methods are traditionally weak. We can quote Singh and Maddala (1976), McDonald and Ransom (1979), McDonald (1984) and more recently Reed and Jorgenson (2004).

However, these new families of densities make the restrictive hypothesis that the income distribution has a single mode, so they cannot detect properly heterogeneity in the sample. A finite mixture of distributions provides a flexible parametric framework for statistical modelling allowing both for flexibility and for the treatment of tail behaviour. The choice of the members of the mixture can be of importance. For instance, Flachaire and Nunez (2002) make use of a finite mixture of normal distributions for modelling log income. Chotikapanich and Griffiths (2008) use a finite mixture of Gamma distributions while Paap and van Dijk (1998) use a finite mixture formed by a Weibull and a truncated normal to model GDP per capita for 120 countries. The log-normal distribution is particularly convenient for modelling the distribution

* Corresponding author. Tel.: +33 491 14 07 45.
E-mail addresses: michel.lubrano@univ-amu.fr (M. Lubrano), abdoul-aziz.ndoye@univ-orleans.fr (A.A.J. Ndoye).

1 Tel.: +33 238 49 24 10.

http://dx.doi.org/10.1016/j.csda.2014.10.009
0167-9473/

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

of incomes in fairly homogeneous sub-populations of the workforce (Aitchison and Brown, 1957), so a finite mixture of log-normal densities can prove to be both a valid and a convenient choice to analyse a heterogeneous population. Validity is based on Ferguson (1983) who has shown that, under some regularity conditions, any distribution can be approximated by a finite mixture of normal distributions. As it can be shown that a finite mixture of normal densities on the log of a variable is the same mathematical object as a finite mixture of log-normal densities, we can assume that Ferguson's result extends to our case. Moreover, in the presence of heavy tails, Mitzenmacher (2004) has shown that a log-normal distribution can exhibit a Pareto tail by expanding its shape parameter, σ^2. The convenience argument is justified as follows. In inequality analysis, decomposition is an important argument as it allows the decomposition of inequality into within-subgroup and between-subgroup components. The additive structure of a mixture model can preserve the decomposability of an inequality index. The log-normal distribution provides analytical formulae for the main inequality indices and their decomposability is preserved by the linear structure of the finite mixture. This decomposability would have disappeared if we had considered a finite mixture of normal densities on the log of the income variable.

We propose a Bayesian approach to inference by Gibbs sampling to model the income distribution using a finite mixture of log-normal densities. We provide statistical inference for some commonly used inequality indices and their decomposition. In recent years, increasing attention has been paid to statistical inference on income inequality measurements. A large number of methods based on asymptotic theory or simulation methods such as bootstrap methods have been proposed in the recent literature with various degrees of complexity (Davidson and Duclos, 2000; Biewen, 2002; Biewen and Jenkins, 2006; Davidson and Flachaire, 2007; Davidson, 2009). Simulation methods are presumed to perform well in small samples. However, Davidson and Flachaire's (2007) Monte Carlo results suggest that bootstrapping commonly used indices of inequality may lead to inference results that are not accurate even in large samples because of possible extreme values in the simulation. Relying on a finite mixture of log-normal densities might solve this kind of problem as the log-normal distribution offers an analytical expression for inequality indices that depends only on the shape parameter of the associated Lorenz curve, σ^2. For decomposable indices, this property extends to mixtures of log-normal distributions. This allows a Rao-Blackwellization procedure that can improve inference results on income inequality indices in finite samples.

We analyse income inequality in the UK over the period 1979–1996 using the Family Expenditure Survey (FES). Since the late 1970s, real income has increased substantially in the UK, but the gap between the poorest and the richest has also increased faster than in any other comparable industrial countries. Our empirical illustration shows that there are large differences in the mean income of each of the mixture components, differences in the structure and in inequality between those components. However, despite these differences, between-group inequality accounts only for a small part of overall inequality changes, around 35%. These results are in line with the explanations put forward in the economic literature (Jenkins, 1996, 2000).

The paper is organized as follows. Section 2 provides elements for Bayesian inference in a finite mixture of log-normal distributions when empirical quantiles are used to elicit prior information and when the marginal likelihood is used to select the optimal number of mixture components. A discussion is led about the influence of this prior on label switching. Section 3 reviews the analytical expressions of commonly used inequality indices, sets out Bayesian inference for these indices and provides the decomposability of the Generalized Entropy index in the framework of mixture models. Section 4 illustrates the approach using the FES data and Section 5 concludes.

2. Finite mixture of log-normal densities

The log-normal density is obtained by considering a transformation of a normal random variable. And this result extends easily to the case of mixtures so that a mixture of log-normal densities is directly obtained by considering the exponential of a random variable that is distributed according to a mixture of normal densities. As a consequence the usual way of estimating a mixture of log-normal densities is to fit a mixture of normal densities on the log of the variable. This is the approach followed for instance by Flachaire and Nunez (2002) and also by many others, justifying the decision of Marin and Robert (2007) to illustrate only the case of mixture of normal densities. However, it is also interesting to consider a mixture of log-normal densities in our case, because this configuration is much more convenient first to elicit prior information and second to decompose inequality indices.

2.1. A Gibbs sampler for a mixture of log-normals

A finite mixture of log-normal densities is a convex combination of k log-normal densities where the density of the jth component is given by $\Lambda(y|\mu_j, \sigma_j^2)$ and where (μ_j, σ_j^2) are the component specific log mean and log variance. If each component is sampled with probability p_j with $\sum p_j = 1$, then the density function of observation y_i, given the vector of parameters (μ, σ^2, p) is:

$$\begin{aligned} f(y_i|\mu, \sigma^2, p) &= \sum_{j=1}^{k} p_j \Lambda(y_i|\mu_j, \sigma_j^2), \\ &= \sum_{j=1}^{k} p_j \frac{1}{y_i\sqrt{2\pi\sigma_j^2}} \exp -\frac{(\log y_i - \mu_j)^2}{2\sigma_j^2}, \end{aligned} \tag{1}$$

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

where $\mu = (\mu_1, \ldots, \mu_k)$, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_k^2)$, $p = (p_1, \ldots, p_k)$. The observed likelihood is obtained as the product of the individual data densities:

$$f(y|\theta) = \prod_{i=1}^{n} \sum_{j=1}^{k} p_j \Lambda(y_i|\mu_j, \sigma_j^2), \tag{2}$$

where θ represents the vector of all the parameters: p, μ, σ^2. As it is very difficult to combine this likelihood function with a natural conjugate prior, it is better to rely on the missing variable representation of Diebolt and Robert (1994). An i.i.d. sample $(y_1, \ldots, y_n)$ generated from (1) may be seen as a collection of sub-samples originating from each of the components $\Lambda(y|\mu_j, \sigma_j^2)$ with unknown size n_j and probability of origin p_j. In an incomplete data problem (the origin of each observation), a missing data representation implies that to each observation y_i is associated a missing variable z_i that indicates from which member of the mixture it originates. Formally, $z_i \in \{1, \ldots, k\}$ follows a multinomial distribution, $z_i|p \sim \mathrm{M}_k(1; p_1, \ldots, p_k)$, while conditionally on z_i, $y_i|z_i, \mu, \sigma^2 \sim \Lambda_j(\cdot|\mu_{z_i}, \sigma_{z_i}^2)$. Knowledge of z_i eliminates the mixing structure present in (2) as the likelihood function is now simplified into:

$$f(y, z|p, \mu, \sigma^2) = \prod_{i=1}^{n} p_{z_i} \Lambda_j(y_i|\mu_{z_i}, \sigma_{z_i}^2). \tag{3}$$

We can define the following sufficient statistics for a particular sample allocation:

$$n_j = \sum_{i=1}^{n} \mathbb{1}(z_i = j), \qquad \bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n} \log y_i \mathbb{1}(z_i = j),$$
$$s_j^2 = \sum_{i=1}^{n} (\log y_i - \bar{y}_j)^2 \mathbb{1}(z_i = j),$$

where $\mathbb{1}(.)$ is the indicator function. Once a sample selection is given, the mixture structure disappears so that conditional on z, each member of the mixture can be treated separately. Defining $\mathcal{I}_j = \{i|z_i = j\}$, we have:

$$\begin{aligned} L(\mu_j, \sigma_j^2|y, z) &= \left(\prod_{\mathcal{I}_j} (y_i)^{-1}\right) (2\pi)^{-n_j/2} \sigma_j^{-n_j} \exp -\frac{1}{2\sigma_j^2} \sum_{\mathcal{I}_j} (\log y_i - \mu_j)^2, \\ &\propto \sigma_j^{-n_j} \exp -\frac{1}{2\sigma_j^2} \sum_{\mathcal{I}_j} (\log y_i - \mu_j)^2, \\ &\propto \sigma_j^{-n_j} \exp -\frac{1}{2\sigma_j^2} \left(s_j^2 + n_j(\mu_j - \bar{y}_j)^2\right). \end{aligned} \tag{4}$$

This conditional likelihood has the same structure as the one we would obtain if we had considered fitting a mixture of Gaussian densities on the logs, which shows as a by-product that for inference it is equivalent to consider a mixture of log-normals or a mixture of normals on the log variable, as neglecting the Jacobian is of no importance.

We can specify natural conjugate priors with a conditional normal prior on μ_j given σ_j^2, and an inverted gamma-2 prior on σ_j^2:

$$\pi(\mu_j|\sigma_j^2) = f_N(\mu_j|\mu_0, \sigma_j^2/n_0) \propto \sigma_j^{-1} \exp -\frac{n_0}{2\sigma_j^2}(\mu_j - \mu_0)^2, \tag{5}$$

$$\pi(\sigma_j^2) = f_{i\gamma}(\sigma_j^2|\nu_0, s_0) \propto \sigma_j^{-(\nu_0+2)} \exp -\frac{s_0}{2\sigma_j^2}. \tag{6}$$

We recall for the unfamiliar reader that the inverted gamma-2 is extensively used in connection with residual variances, see for instance Bauwens et al. (1999). It is defined as

$$f_{i\gamma}(\sigma^2|\nu, s) \propto (\sigma^2)^{-\frac{1}{2}(\nu+2)} \exp -\frac{s}{2\sigma^2}.$$

See the appendix of Bauwens et al. (1999) for the constant of integration and moments.

The conditional posterior of $\mu_j|\sigma_j^2$ is normal with:

$$\begin{aligned} \pi(\mu_j|\sigma_j^2, y, z) &\propto \sigma_j^{-1} \exp -\frac{1}{2\sigma_j^2} \left((n_0\mu_0 + n_j\bar{y}_j)/n_{*j}\right), \\ &= f_N(\mu_j|\mu_{*j}, \sigma_j^2/n_{*j}), \end{aligned} \tag{7}$$

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

where $n_{*j} = n_0 + n_j$ and $\mu_{*j} = (n_0\mu_0 + n_j\bar{y}_j)/n_{*j}$. The conditional posterior density of σ^2 is an inverted gamma-2:

$$\pi(\sigma_j^2|y,z) \propto \sigma_j^{-(n_j+\nu_0+2)} \exp -\frac{1}{2\sigma_j^2}\left(s_0 + s_j^2 + \frac{n_0 n_j}{n_0+n_j}(\mu_0 - \bar{y}_j)^2\right),$$
$$= f_{i\gamma}(\sigma_j^2|\nu_{*j}, s_{*j}). \tag{8}$$

where $\nu_{*j} = \nu_0 + n_j$ and $s_{*j} = s_0 + s_j^2 + \frac{n_0 n_j}{n_0+n_j}(\mu_0 - \bar{y}_j)^2$.

Let us now turn to the case of p. We specify a Dirichlet prior on p:

$$\pi(p) = f_{Dir}(\gamma_1^0, \ldots, \gamma_k^0) \propto \prod_{j=1}^{k} p_j^{\gamma_j^0 - 1}. \tag{9}$$

The joint posterior probability density function of p_j conditional on z is also Dirichlet with:

$$\pi(p|y,z) = f_{Dir}(\gamma_1^0 + n_1, \ldots, \gamma_k^0 + n_k) \propto \prod_{j=1}^{k} p_j^{\gamma_j^0 + n_j - 1}. \tag{10}$$

It combines the prior value of γ_j^0 with the size of the sub-sample allocated to the jth member of the mixture.

Given values or draws of p, μ and σ^2, we can define the posterior probability that $z_i = j$:

$$\Pr(z_i = j|y,\theta) = \frac{p_j \Lambda(y_i|\mu_j, \sigma_j^2)}{\sum_j p_j \Lambda(y_i|\mu_j, \sigma_j^2)}. \tag{11}$$

As pointed out by one of the referees, the Jacobian of the transformation $x = \log y_i$ can be factorized both in the numerator and denominator of (11) so that $\Pr(z_i = j)$ is strictly identical in a mixture of log-normals on y and in a mixture of normals on $x = \log y$ so that here again the Jacobian can be ignored.

Finally, the posterior predictive density $\widehat{f}(y_i)$ is defined as follows:

$$\widehat{f}(y_i) = \mathrm{E}_\theta[f(y_i|\theta)|y],$$
$$= \mathrm{E}_\theta[\sum_{j=1}^{k} p_j \Lambda(y_i|\mu_j, \sigma_j^2)|y]. \tag{12}$$

A Gibbs sampler, as detailed in Diebolt and Robert (1994), Frühwirth-Schnatter (2006) or Marin and Robert (2007) alternatively generates the z given a sample y and a previous draw of the parameters (p, μ, σ^2) and then generates the parameters (p, μ, σ^2) given a sample y and the previous draws for the z.

Gibbs sampler algorithm for a mixture of log-normal distributions

(1) Fix k the dimension of the finite mixture, m the number of draws, set the vectors of starting values $p^{(0)}$, $\mu^{(0)}$, $\sigma^{2(0)}$.
(2) For $l = 1, \ldots, m$
   (a) Generate a sample separation $z_i^{(l)}$ $(i = 1, \ldots, n,\ j = 1, \ldots, k)$ with probability
   $$\Pr\left(z_i^{(l)} = j|p_j^{(l-1)}, \mu_j^{(l-1)}, \sigma_j^{2(l-1)}, y_i\right) \propto p_j^{(l-1)} \Lambda\left(y_i|\mu_j^{(l-1)}, \sigma_j^{2(l-1)}\right).$$
   (b) Compute as a by-product the statistics:
   $$n_j^{(l)} = \sum_{i=1}^{n} \mathbb{1}(z_i^{(l)} = j),$$
   $$\bar{y}_j^{(l)} = \sum_{i=1}^{n} \mathbb{1}(z_i^{(l)} = j) y_i / n_j^{(l)},$$
   $$s_j^{(l)} = \sum_{i=1}^{n} \mathbb{1}(z_i^{(l)} = j)(y_i - \bar{y}_j^{(l)})^2.$$
   (c) Generate the vector $p^{(l)}$ from the posterior Dirichlet density
   $$\mathrm{Dir}\left(\gamma_1^0 + n_1^{(l)}, \ldots, \gamma_k^0 + n_k^{(l)}\right).$$
   (d) Generate $\sigma_j^{2(l)}$ from the inverted gamma-2 posterior density
   $$f_{i\gamma}(\sigma_j^{2(l)}|\nu_{*j}, s_{*j}).$$
   (e) Generate $\mu_j^{(l)}$ conditionally on the draw $\sigma_j^{2(l)}$ from the conditional normal posterior density
   $$f_N(\mu_j|\mu_{*j}, \sigma_j^2/n_{*j}).$$

(f) Evaluate the conditional posterior predictive density

$$\widehat{f}^{(l)}(y|z_i^{(l)}) = \sum_{j=1}^{k} p_i \Lambda(y|\mu_j^{(l)}, \sigma_j^{2(l)}).$$

(3) Marginalize over the draws the conditional posterior predictive density.
(4) Marginalize over the draws the conditional posterior moments of the parameters.

2.2. Reducing label switching using an informative prior

All of the posterior densities detailed in the previous section are well defined only when there is enough prior information. Prior information is always needed for Bayesian inference in finite mixtures. For instance, a small value of n_j has to be compensated by a greater value of γ_j^0 so that $\gamma_j^0 + n_j - 1 > 0$ in (10). In practice this means that we must specify at least $\gamma_j^0 > 1$. The conditional posterior densities (7) and (8) are well conditioned provided $n_0 > 0,\ s_0 > 0$ and $\nu_0 > 0$.

The posterior densities exist as soon as we provide the same minimal prior information to the k mixture components, information which can be based on sample moments as described for instance in Marin and Robert (2007). This is equivalent to saying a priori that the sample could be represented by a unique log-normal density. For instance, μ_0 is set to the log sample mean and $s_0/(\nu_0 - 2)$ is set proportional to the log sample variance, and this for all the members of the mixture. This type of information is sufficient to avoid the breakdown of the Gibbs sampler. But it is not sufficient to avoid label switching. Label switching is of no importance when running the Gibbs sampler. However, it blurs the allocation of an observation to a particular cluster and precludes reporting posterior moments for the parameters. In their seminal paper, Diebolt and Robert (1994), propose a reordering on one of the parameters corresponding for instance to imposing $\mu_1 < \cdots < \mu_k$. Jasra et al. (2005) report that this is equivalent to changing the initial prior $\pi(\theta)$ into:

$$\pi_n(\theta) = k!\pi(\theta)\mathbb{1}(\theta \in C),$$

where C is the constraint. The original symmetry of a mixture is broken and the sample has to be represented a priori by several log-normal densities instead of a single one. This deterministic prior was shown to be too strong in the literature (see the papers of Celeux et al., 2000, Stephens, 2000b, Frühwirth-Schnatter, 2001 or Jasra et al., 2005 for a general discussion on label switching). The final result is not invariant to the choice of the parameter to order (see for instance the experiments in Marin et al., 2005). We would like here to take advantage of the specific distribution we want to analyse, the income distribution, in order to specify an informative prior in a probabilistic way. Our aim is to identify different groups in the population, namely the poor, the middle class and the rich, in order to be able to decompose an inequality index. The probability of label switching can be reduced (but not totally eliminated) if the prior is informative enough. An informative prior will identify each member in a probabilistic manner. Calibrating the degree of precision of the prior is certainly a difficult task and we shall report variants in the empirical application.

The poor category is defined as those who have an income below 60% of the median income (or 50% of the mean income) and the rich by those who have an income greater than a given quantile, say $q_{0.90}$. So sample quantiles may provide information on the location parameters of the mixture as with a log-normal density $\Lambda(\mu, \sigma^2)$, the median is equal to $\exp(\mu)$ while the mean is $\exp(\mu + \sigma^2/2)$. Eliciting a prior value for σ^2 can be more simple as this parameter is scale free in the log-normal. We know that in the log-normal case the Gini index is equal to $2\Phi(\sigma/\sqrt{2}) - 1$, with $\Phi(.)$ being the cumulative distribution of the standard normal distribution. So $\sigma^2 = 1$ would correspond to a Gini of 0.383, while $\sigma^2 = 0.5$ would correspond to $G = 0.197$. However, the prior on σ^2 can also be estimated so as to verify that the implied prior means of the income variable for the different groups are ordered and that income groups in the right tail of the asymmetric income distribution correspond to larger values of σ_j^2 (a property of the log-normal density).

Let us now detail how the procedure could work. For each mixture component, we have to specify five hyper-parameters, (s_0, ν_0), (μ_0, n_0) and γ_0 appearing in:

$$\begin{array}{lll}
\pi(\sigma_j^2): & \mathrm{E}(\sigma_j^2) = s_0/(\nu_0 - 2), & \\
\pi(\mu_j|\sigma_j^2): & \mathrm{E}(\mu_j) = \mu_0, & \mathrm{Var}(\mu_j|\sigma_j^2) = s_0/(\nu_0 - 2)/n_0, \\
\pi(p_j): & \mathrm{E}(p_j) = \gamma_{0j}/\sum_j \gamma_{0j}. &
\end{array}$$

It is possible to fix ν_0, the prior degree of freedom of the inverted gamma-2 prior (6) and n_0, the prior precision of the normal prior (5) at predefined values such that for instance $\nu_0 = 5$ and $n_0 = 1$. The prior ordering of the mean income of each group corresponds to the following set of constraints:

$$\mu_{01} + \frac{s_{01}}{\nu_{01} - 2} < \mu_{02} + \frac{s_{02}}{\nu_{02} - 2} < \cdots < \mu_{0k} + \frac{s_{0k}}{\nu_{0k} - 2},$$

that are easy to check. Finally, an identical prior can be chosen for the parameters of the Dirichlet prior (9) on the p_i, provided it is not too informative in order to avoid conflicts of information.

To be coherent with this prior information, the posterior draws should verify:

$$\mu_1^{(l)} + \frac{\sigma_1^{2(l)}}{2} < \mu_2^{(l)} + \frac{\sigma_2^{2(l)}}{2} < \cdots < \mu_k^{(l)} + \frac{\sigma_k^{2(l)}}{2}.$$

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

The number of times this constraint is violated in the MCMC output can be taken as an indication for the importance of remaining label switching. But we do not say of course that the labels in the MCMC output have to be permuted according to this scheme.

2.3. Selecting the optimal number of mixture components

Choosing the number of components in a mixture can be seen as a choice of a model problem. Varying the number of components k is equivalent to defining different models, like choosing the number of lags in a dynamic model. However, from a different point of view, k could be seen as a parameter for which a posterior density could be derived. Such an approach is illustrated for instance by the reversible jump algorithm of Green (1995). See also the birth-and-death approach of Stephens (2000a). Both methods are discussed in Marin et al. (2005).

In a Bayesian framework, model choice relies on the evaluation of the marginal likelihood of the different models and on Bayes factors. The evaluation of a marginal likelihood is a difficult task because it means integrating the likelihood function with respect to the prior and this integral does not exist if the prior is non-informative. However, even if the prior is informative, the result is very often numerically unstable, so other ways have been looked at in the literature (see Kass and Raftery, 1995 for a survey). We shall illustrate in three different methods which make use of the MCMC output. They correspond either to an information criterion which penalizes a measure of fit by a measure of complexity or to an evaluation of a MCMC predictive density. The problem is made more complex for mixtures of densities because the parameters of interest and the dimension of the model are not precisely defined. The BIC criterion of Schwarz (1978), which is based on asymptotic expansions, simply considers the number of initial parameters equal to $3k - 1$ and corresponds to:

$$BIC(k) = -2\sum_i \log f(y_i|\tilde{\theta}) - \frac{3k-1}{2}\log(n). \tag{13}$$

The BIC is evaluated at $\tilde{\theta}$ which represents a particular choice for a point value and is discussed later. However, considering $3k - 1$ for measuring complexity is not obvious as illustrated by the existence of two representations for the likelihood function: the observed likelihood $f(y|\theta)$ presented in (2) and the complete likelihood $f(y, z|\theta)$ detailed in (3). The deviance information criteria (DIC) of Spiegelhalter et al. (2002) is an attempt to solve the question of the measurement of model complexity. Roughly speaking, a deviance information criteria is based on the deviance $D(\theta) = -2\log f(\theta|y)$. It compares the MCMC expectation of the deviance $\overline{D(\theta)}$, to the evaluation of the deviance at a single point $D(\tilde{\theta})$. The complexity of the model is defined as the scale independent difference $p_D = \overline{D(\theta)} - D(\tilde{\theta})$ while the DIC is defined as the sum of $\overline{D(\theta)}$ and the complexity measure p_D. The use of a DIC is not evident in the context of missing data models. This question was extensively discussed in Celeux et al. (2006) who have proposed several possible choices for mixtures. We have selected two of them, keeping their notation. When using the observed likelihood, a convenient formulation of the DIC is their DIC_2 or DIC_3:

$$DIC_2 = -4\mathrm{E}_\theta[\log f(y|\theta)|y] + 2\log f(y|\tilde{\theta}(y)), \tag{14}$$

$$DIC_3 = -4\mathrm{E}_\theta[\log f(y|\theta)|y] + 2\log \hat{f}(y). \tag{15}$$

The expectation in the first term is easily approximated by $\frac{1}{m}\sum_{l=1}^{m}\log f(y|\theta^{(l)})$. The second term in DIC_2 is easily computed once $\tilde{\theta}(y)$ is chosen. The last term in DIC_3 corresponds to the MCMC predictive density and is evaluated as $\prod_{i=1}^{n}\mathrm{E}_\theta[f(y_i|\theta)|y]$ from the MCMC output. A total of seven different variants were proposed and evaluated in Celeux et al. (2006). Some of them are shown to provide poor results in the context of finite mixtures. One of their preferred choice is DIC_3 (together with DIC_4 not detailed here). However, they point out that "*DIC*s can be seen as a Bayesian version of *AIC* and [...] they may under-penalize model complexity". Consequently, a properly computed BIC is likely to select a more parsimonious model.

The choice of $\tilde{\theta}(y)$ is of particular importance in the context of mixtures. Because of possible label switching, the different components of the mixture are hard to identify so choosing the posterior mean for $\tilde{\theta}(y)$ can lead taking the average of several distant modes and has to be avoided. A much better choice for $\tilde{\theta}(y)$ is $\arg\max_\theta f(\theta|y)$, the maximum a posteriori estimator that can be derived from the MCMC output. This remark remains valid also for the BIC criterion and we shall adopt it.

Chib (1995) has developed another method to evaluate the MCMC predictive density using a MCMC output. In Bayes' theorem, the marginal likelihood appears as the integrating constant of the posterior density:

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{m(y)},$$

and is valid for any value of θ. If we select a particular θ^* and take logs, this formula can be rearranged so as to obtain:

$$\log\hat{m}(y) = \log f(y|\theta^*) + \log\pi(\theta^*) - \log\hat{\pi}(\theta^*|y). \tag{16}$$

Once a value for θ^* has been chosen, it is straightforward to compute the value of the two first elements of the right hand side of (16). The computation of the last element in (16) is of course more problematic as we do not know the analytical expression of the posterior density and its integrating constants, unless we adopt the missing variable representation of the likelihood function. We first have

$$\pi(\theta_j|y) = \int \pi(\mu_j, \sigma_j^2|y, z)\pi(p_j|y, z)p(z|y)dz.$$

Table 1
Inequality indices for the log-normal distribution.

Inequality index	General expression	Expression of the log-normal
I_G	$\frac{1}{\mu}\int_0^\infty F(y)(1-F(y))dy$	$I_G(\sigma^2) = 2\Phi(\sigma/\sqrt{2}) - 1$
I_{GE}^α	$\frac{1}{\alpha^2-\alpha}\int\left[\left(\frac{y}{\mu(F)}\right)^\alpha - 1\right]f(y)dy$	$I_{GE}^\alpha(\sigma^2) = \frac{\exp((\alpha^2-\alpha)\sigma^2/2)-1}{\alpha^2-\alpha}$
I_A^ϵ	$1-\left(\int\left(\frac{y}{\mu(F)}\right)^{1-\epsilon}f(y)dy\right)^{\frac{1}{1-\epsilon}}$	$I_A^\epsilon(\sigma^2) = 1-\exp\left(-\frac{1}{2\epsilon\sigma^2}\right)$

I_G is the Gini index, I_{GE}^α is the Generalized Entropy index, I_A^ϵ is the Atkinson index.

We know from above that our conditional posterior $\pi(\mu_j, \sigma_j^2|y, z)$ is a normal-inverted-gamma-2 density while $\pi(p_j|y, z)$ is Dirichlet. We know all the integrating constants of these densities (see Appendix A of Bauwens et al., 1999). As a by-product of the Gibbs output we have draws of the z so that we can approximate the log of these posterior densities using:

$$\hat{\pi}(\mu_j^*, \sigma_j^{2*}|y) \simeq \frac{1}{m}\sum_{l=1}^{m}\pi(\mu_j^*, \sigma_j^{2*}|y, z^{(l)}) \quad \text{and} \quad \hat{\pi}(p_j^*|y) \simeq \frac{1}{m}\sum_{l=1}^{m}\pi(p_j^*|y, z^{(l)}).$$

The selection $\theta_j^* = (\mu_j^*, \sigma_j^{2*}, p_j^*)$ should not be a problem as in theory any value can be chosen. Here again, we can choose $\theta^* = \tilde{\theta}$, the maximum a posteriori estimator. But that choice would imply running the MCMC simulator twice (the first time for finding θ_j^*, the second time for computing the predictive density at that point). As in a mixture context we have to be informative on all the parameters, we can decide to select θ_j^* as being the prior mean. So we can control exactly at which point $\log \hat{m}(y)$ is evaluated when comparing models.

The application of Chib's method in the context of a finite mixture of distributions is problematic because of the label switching problem. As explained in Frühwirth-Schnatter (2004), the computation of the unconstrained marginal likelihood should be done over all the $k!$ possible sample separations. If we are sure that there is no label switching in a MCMC output, the expected bias of Chib's method is known to be equal to $-\log k!$. In this case, we simply correct for the bias by adding $\log k!$ to the estimated marginal likelihood. In the intermediate case where in a MCMC output, we have a moderate label switching, the bias correction is no longer possible. Nevertheless, if when comparing two models, in the ignorance of the importance of the label switching, the difference between the two log marginal likelihoods is greater than $\log k!$, the bias of the method is no longer of any importance as it will not change the decision (to fix ideas, $\log 3! = 1.79$, $\log 4! = 3.18$ and $\log 5! = 4.79$).

The same kind of remark can be made for the *DIC*. As soon as a particular version of the DIC involves an expectation over θ, it is made sensitive to label switching, which can explain the reserved results reported in the empirical illustrations of Celeux et al. (2006). On the contrary the *BIC* criterion does not seem to be sensitive to label switching as it relies only on the value $\tilde{\theta}(y)$, and we choose to select the maximum a posteriori estimator for it. Note also that Frühwirth-Schnatter (2006, p. 117) reports that the BIC criterion is consistent for selecting the number of components provided the family of component densities is correctly specified.

3. Bayesian inference on inequality measurements

Finite mixtures of log-normals prove to be very useful when analysing inequality and inequality decomposition. First of all, under suitable prior information, we can try to identify each member of the mixture with a social group according to its income level. Second, using decomposable indices, we can analyse inequality inside each group and between the groups identified by the mixture. Finally, the possibility of Rao-Blackwellization can help to improve the precision of inference results for these indices, knowing that a finite mixture of log-normals provides conditional analytical results for conditional moments of these indices.

3.1. Inequality measurements and the log-normal distribution

Cowell (1995) provides the different analytical expressions for commonly used inequality indices based on the log-normal distribution that we reproduce in Table 1. Each measure depends on the single shape parameter σ^2 of the log-normal. $\Phi(.)$ is the cumulative distribution of the standard normal distribution and $\mu(F)$ is the income average of the population considered having distribution F. We consider $\alpha \in (-\infty, +\infty)$ as the parameter that captures the sensitivity of a specific Generalized Entropy (GE) index to particular parts of the distribution: for large and positive values of α, the index is sensitive to changes in the upper tail of the distribution; for α negative, the index is sensitive to changes in the lower tail of the distribution (Cowell, 1995; Bourguignon, 1979). In empirical work, the range of values for α is typically restricted to $[-1, 2]$ because, otherwise, estimates may be unduly influenced by a small number of very small or very high incomes (see for instance Shorrocks, 1980).

The parameter $\epsilon \geq 0$ characterizes (relative) inequality aversion for the Atkinson index, inequality aversion being an increasing function of ϵ. The Atkinson index may be viewed as a particular case of the GE index with $\alpha \leq 1$ and $\epsilon = 1 - \alpha$.

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

Measures which are ordinally equivalent to the GE class include a number of pragmatic indexes such as the mean logarithmic deviation index ($I_{MLD} = \lim_{\alpha\to 0} I_{GE}^{\alpha}$), Theil's index ($I_{Theil} = \lim_{\alpha\to 1} I_{GE}^{\alpha}$) and the coefficient of variation ($1/2I_{CV}^2 = \lim_{\alpha\to 2} I_{GE}^{\alpha}$). For the log-normal distribution, the MLD and the Theil index become the same and are both equal to $\sigma^2/2$.

As the GE indices are decomposable, we can recover an analytical expression when F is a finite mixture of log-normals. This decomposability will allow for a Rao-Blackwellization from the Gibbs output when computing these indices and should thus provide us with a more precise MCMC evaluation of their posterior standard deviations.

3.2. Properties of mixture models

Mixture models have nice properties that will be of direct interest for our purpose. Those properties are directly related to the linearity of the model. In any finite mixture, the overall cumulative distribution is obtained as the weighted sum of the individual cumulative distributions so that in our case we have:

$$F(x) = \sum_{j=1}^{k} p_j F_j(x|\mu_j, \sigma_j^2).$$

The first moment $\mu(F)$ of X is obtained as a linear combination of the first moment of each member of the mixture:

$$\mu(F) = \sum_{j=1}^{k} p_j \mu(F)_j.$$

That property extends to the un-centred higher moments.

We can use directly these properties in order to derive the expression of the Gini index for a mixture of log-normals. A Gini index can be written as a function of the overall cumulative distribution, using the integral expression given in Table 1:

$$I_G(\mu, \sigma^2, p) = \frac{1}{\mu(F)} \int_0^{\infty} F(x)(1 - F(x))\, dx.$$

For a mixture of k elements, we have:

$$I_G(\mu, \sigma^2, p) = \frac{1}{\sum_{j=1}^{k} p_j \mu(F)_j} \int_0^{\infty} \sum_{j=1}^{k} p_j F_j(x) \left(1 - \sum_{j=1}^{k} p_j F_j(x)\right) dx.$$

As the cumulative of the log-normal is $F_j(x) = \Phi(\frac{\log x - \mu_j}{\sigma_j})$, the Gini index can be obtained as the result of a simple numerical integral ($\Phi(.)$ being directly available in any numerical package). However, Φ can be also approximated by polynomial series or hyper-geometric functions as suggested in Abadir (1999).

This integral has to be evaluated for every draw of the MCMC experiment. We thus get m evaluations of the Gini index. Summing over all the draws, we get an estimate for the mean index:

$$\hat{I}_G = \frac{1}{m} \sum_{t=1}^{m} I_G(\mu^{(l)}, \sigma^{2(l)}, p^{(l)}).$$

The standard deviation can be obtained in a similar way by summing the squares:

$$\hat{I}_G^2 = \frac{1}{m} \sum_{t=1}^{m} I_G(\mu^{(l)}, \sigma^{2(l)}, p^{(l)})^2,$$

so that the small sample variance is obtained as $\hat{I}_G^2 - (\hat{I}_G)^2$.

For decomposable indices, it is possible to go a step further on as decomposability implies that the overall index can be expressed as a weighted sum of individual indices (plus a remainder) as we shall now see. This is the way to get the usual Rao-Blackwellization (average of a conditional expectation).

3.3. Index decomposability and mixture models

Definition 1 (*Shorrocks, 1980*). Given a population of any size $n \geq 2$ and a partition into k non-empty subgroups, an inequality index $I(y, n)$ is decomposable if there exists a set of coefficients $\tau_j^k(\mu, n)$ such that:

$$I(y, n) = \sum_{j=1}^{k} \tau_j^k I(y^j; n_j) + B, \quad y = (y^1, \dots, y^k),$$

where $\mu = (\mu_1, \dots, \mu_k)$ is the vector of subgroup means, $\tau_j(\mu, n)$ is the weight attached to subgroup j in a decomposition into k subgroups and B is the between-group term assumed to be independent of inequality within the individual subgroups, $B = I(\mu_1 u_{n_1}, \dots, \mu_k u_{n_k})$ where u_{n_i} represents the unit vector $(1, 1, \dots, 1)$ with n components.

Only few indices are decomposable. The most famous counter-example is the Gini index, the decomposition of which has, in most cases, a remainder term which is not directly interpretable. The most widely used decomposable index is the GE index. The Atkinson index is only indirectly decomposable. Atkinson and GE indices have different cardinalization functions but they are ordinally equivalent for cases $\alpha \leq 1$ and $\epsilon = 1 - \alpha$ since:

$$I_A^{\epsilon}(F) = 1 - \frac{1}{\mu(F)}\left[(\alpha^2-\alpha) I_{GE}^{\alpha}(F) + 1\right]^{\frac{1}{\alpha}}.$$

Let us now study in detail the case of the GE index. For a given mixture model with k components, the class of GE indices can be expressed as:

$$\begin{aligned} I_{GE}^{\alpha} &= \frac{1}{\alpha^2-\alpha}\int\left[\left(\frac{y}{\sum_{j=1}^{k} p_j\mu_j}\right)^{\alpha} - 1\right]\sum_{j=1}^{k} p_j f_j(y)dy, \\ &= \sum_{j=1}^{k} p_j \frac{1}{\alpha^2-\alpha}\int\left[\left(\frac{y\mu_j}{\mu_j\sum_{j=1}^{k} p_j\mu_j}\right)^{\alpha} - 1\right] f_j(y)dy, \\ &= \sum_{j=1}^{k} p_j\left(\frac{\mu_j}{\sum_{j=1}^{k} p_j\mu_j}\right)^{\alpha}\frac{1}{\alpha^2-\alpha}\int\left[\left(\frac{y}{\mu_j}\right)^{\alpha} - 1\right] f_j(y)dy + \frac{1}{\alpha^2-\alpha}\left[\sum_{j=1}^{k} p_j\left(\frac{\mu_j}{\sum_{j=1}^{k} p_j\mu_j}\right)^{\alpha} - 1\right]. \end{aligned}$$

If $\tau_j = p_j\mu_j / \sum_{j=1}^{k} p_j\mu_j$ and I_{GE}^{j} denotes the GE family index with parameter α for the group j, then we have the following decomposition:

$$I_{GE}^{\alpha} = \underbrace{\sum_{j=1}^{k} p_j^{1-\alpha}\tau_j^{\alpha} I_{GE}^{j}}_{\text{within}} + \underbrace{\frac{1}{\alpha^2-\alpha}\left(\sum_{j=1}^{k} p_j^{1-\alpha}\tau_j^{\alpha} - 1\right)}_{\text{between}}. \tag{17}$$

Within the GE family, the Theil and the MLD are the only zero homogeneous decomposable measures such that the weights of the within-group-inequalities in the total inequality sum to a constant (Bourguignon, 1979). They are expressed from a mixture point of view as:

$$I_{Theil} = \sum_{j=1}^{k} \tau_j I_{Theil}^{j} + \sum_{j=1}^{k} \tau_j \log\left(\frac{\tau_j}{p_j}\right),$$

$$I_{MLD} = \sum_{j=1}^{k} p_j I_{MLD}^{j} + \sum_{j=1}^{k} p_j \log\left(\frac{p_j}{\tau_j}\right).$$

When applied to mixtures of log-normal distributions, the individual indices are equal: $I_{Theil}^{j} = I_{MLD}^{j} = \sigma_j^2/2$. But the Theil and MLD indices have distinct expressions for the total mixture, simply because they have different weights: τ_j in first case, p_j in the second. The τ_j weights of the Theil index are given by $\tau_j = p_j\exp(\mu_j + \sigma_j^2/2)/\sum p_j \exp(\mu_j + \sigma_j^2/2)$.

3.4. Rao-Blackwellization from the Gibbs output

We consider $\left(\mu^{(l)}, \sigma^{2(l)}, p^{(l)}\right)$ for $l = 1, \ldots, m$ as the Gibbs output obtained by m MCMC-generated draws for a k-components mixture of log-normals. We can obtain posterior means and standard deviations for the GE inequality indices by Rao-Blackwellization of the Gibbs output using (17). So the posterior mean and posterior variance of the GE index are obtained from the following sums, without any further numerical procedure:

$$\hat{I}_{GE}^{\alpha} = \frac{1}{m}\sum_{l=1}^{m} I_{GE}^{\alpha}(\mu^{(l)}, \sigma^{2(l)}, p^{(l)}),$$

$$\hat{\sigma}^2(I_{GE}^{\alpha}) = \frac{1}{m}\sum_{l=1}^{m}(I_{GE}^{\alpha}(\mu^{(l)}, \sigma^{2(l)}, p^{(l)}))^2 - (\hat{I}_{GE}^{\alpha})^2.$$

We could not obtain such a direct result for the Gini index in Section 3.2, but only an approximation relying on an extra numerical integration.

Table 2
Quantile based prior information.

	μ_{01}	μ_{02}	μ_{03}	μ_{04}	Mean log
1979	3.96	4.31	4.64	5.06	4.31
1988	4.05	4.47	4.85	5.37	4.48
1992	4.09	4.52	4.92	5.46	4.52
1996	4.19	4.55	4.92	5.44	4.57

The μ_{0j} correspond to the log quantiles of level 0.25, 0.50, 0.75, 0.95.

4. Application: modelling the UK income distribution

Since the late 1970s, real income has increased substantially in the UK, but the gap between the poorest and the richest has also increased faster than in any other comparable industrial countries. We will provide a decomposition analysis of this increase in inequality using the Family Expenditure Survey (FES).

4.1. Data

The Family Expenditure Survey (FES) is a continuous sample survey of the UK population living in households. Our sample covers four waves of the survey: 1979, 1988, 1992 and 1996 which contain the period when Mrs. Thatcher was the prime minister (1979–1990). The data correspond to household disposable income (i.e. post-tax and after transfer incomes), but before housing costs. Household disposable income is modelled on the basis of a pseudo-panel data set and is equivalized by the McClements adult-equivalence scale and deflated by the corresponding relative consumer price index. These data are the same as those used by Flachaire and Nunez (2002), except that we deflate by a consumption price index, while they take the observations in deviations from the mean. We have 6230 observations in 1979, 6456 in 1988, 6597 in 1992, 6043 in 1996.

4.2. Model selection

We provide prior information that serves both for inference and for model selection. We have a strong prior for a mixture with three members because we want to interpret each member as representative of one of the social classes: poor, middle, and rich. But for model selection purposes, we consider the possibility of more members. We selected four quantiles of the density of log incomes 0.25, 0.50, 0.75, 0.95 and also 0.99 which is needed in one case. These values will serve to define prior expectations for the μ_j. We chose $n_0 = 1$ as prior precision. For the prior on σ^2, we selected a prior mean of 0.50, common for all the samples as σ^2 is scale free in the log-normal distribution. We chose a common value of 5 for ν_0. For γ_0, we selected the value of 5 for all the members of the mixture, which is a rather soft prior when compared to the sample size. Table 2 summarizes the different sample quantiles and means of the logs.

We used the Gibbs algorithm described above to generate 10,000 draws from the posterior density in order to select the optimal number of mixture components. The chain was run with 10 000 draws plus 1000 draws for warming. We computed various indicators: the BIC (evaluated at the maximum a posteriori estimator), the marginal likelihood using Chib's method (computed at the prior mean), and two deviance information criteria, DIC_2 and DIC_3. Most of the time, BIC and Chib were in agreement to select a model with 3 components. The deviance information criterion provided the same answer only for 1988. Otherwise it continues to decrease as an inverse function of k. With those data sets, we cannot get a general and clear answer for selecting mixture components using a DIC, corroborating the extensive results of Celeux et al. (2006). Chib's method does not seem to be too influenced by label switching in this example, because adding $\log k!$ to the log of the marginal likelihood does not change the ordering. Finally, we undertook a sensitivity analysis. The last block of Table 3 is devoted to analysing a variant consisting in adopting a unique prior mean for μ equal to the sample mean of the logs, and this only for the year 1988. We get an identical conclusion for Chib's predictive and BIC. However, we can no longer get a conclusion with the DIC. 1988 was the only period for which the DIC gave a similar answer. Just by changing the prior, we lose this unique result, which shows the fragility of the DIC in the case of mixtures. We can however conclude that the best approximating model is a finite mixture of three log-normal densities.

Using the same data, but fitting a mixture of normals on the log variable, Flachaire and Nunez (2002) found a much greater number of components for their mixture, five to seven. But they use a slightly different model where the probability that an observation belongs to a particular group is not determined by (11), but by an additional model including exogenous variables related to household characteristics and job status. This is useful for identifying the groups. We use instead probabilistic information to characterize three groups by income levels.

4.3. Posterior results

The first posterior results we obtained under the previous prior presented some label switching because the groups were not identified in the proper order required by the prior on μ. In order to reduce label switching, we modified our initial prior information and augmented the asymmetric informative prior on μ_j with an asymmetric prior information on σ_j^2, selecting

Table 3
Model selection criteria using an asymmetric prior on μ_i.

		$k=2$	$k=3$	$k=4$	$k=5$
1979	Chib	−20 410.12	**−19 724.48**	−20 743.39	
	BIC	61 637.06	**61 387.91**	61 404.63	
	DIC_2	61 613.89	61 352.58	61 342.68	
	DIC_3	61 608.36	61 344.11	61 334.38	
1988	Chib	−25 395.35	**−23 457.88**	−24 840.93	
	BIC	68 106.44	**67 920.84**	67 943.77	
	DIC_2	68 087.26	**67 883.53**	67 884.05	
	DIC_3	68 080.78	**67 875.75**	67 875.75	
1992	Chib	−25 631.59	**−24 862.64**	−24 954.15	−26 716.43
	BIC	71 109.3	70 918.72	**70 886.30**	70 912.24
	DIC_2	71 078.64	70 885.89	70 826.99	70 826.62
	DIC_3	71 078.54	70 881.15	70 817.62	70 817.50
1996	Chib	−22 682.85	**−22 534.55**	−23 451.64	
	BIC	64 570.97	**64 453.10**	64 459.61	
	DIC_2	64 543.54	64 415.81	64 405.31	
	DIC_3	64 540.59	64 408.29	64 393.96	
	Using an identical prior on μ_i centred on the sample mean				
1988	Chib	−23 886.05	**−23 201.25**	−23 906.44	−23 907.48
	BIC	68 233.32	**67 920.70**	67 947.84	67 970.95
	DIC_2	68 204.8	67 889.38	67 886.35	67 885.05
	DIC_3	68 202.66	67 878.52	67 879.2	67 876.92

Table 4
Posterior moments of mixture parameters.

	μ_1	μ_2	μ_3	σ_1^2	σ_2^2	σ_3^2	p_1	p_2	p_3
1979	3.90	4.49	3.82	0.062	0.156	1.49	0.290	0.700	0.010
	(0.015)	(0.018)	(0.339)	(0.005)	(0.007)	(0.245)	(0.025)	(0.026)	(0.003)
1988	3.99	4.68	4.46	0.068	0.208	1.52	0.285	0.687	0.029
	(0.015)	(0.019)	(0.125)	(0.005)	(0.010)	(0.217)	(0.022)	(0.024)	(0.007)
1992	4.02	4.75	4.12	0.084	0.221	2.88	0.291	0.684	0.025
	(0.017)	(0.025)	(0.172)	(0.007)	(0.013)	(0.367)	(0.029)	(0.029)	(0.004)
1996	4.20	4.77	3.98	0.109	0.232	2.47	0.344	0.645	0.012
	(0.029)	(0.034)	(0.589)	(0.011)	(0.014)	(0.422)	(0.048)	(0.048)	(0.004)

the prior expectations equal to (0.25, 0.50, 1.50) with prior degrees of freedom equal to 50. This new prior manages to reduce label switching because now the groups do correspond to their prior order. In order to appraise the convergence of the Gibbs sampler, we use CUMSUM graphs, as displayed in Figs. 1–4. CUMSUM graphs were first proposed by Yu and Mykland (1998) as a simple diagnostic method in order to assess the convergence of MCMC chains. The idea is to plot the evolution of a partial moment. If x_j is the jth MCMC draw, then $\bar{x}$ represents the MCMC estimate of the mean and $\hat{\sigma}_x$ the MCMC estimate of the standard deviation, both over the complete MCMC set. The (standardized) partial moment computed over the first i draws is $c_i = \sum_{j=1}^{i}(x_j/i - \bar{x})/\hat{\sigma}_x$. The standardized CUMSUM graph is formed by plotting c_i against i for $i = 1, \ldots, m$ where m is the total number of draws. This is a standardized version of the diagnostic procedure as proposed and detailed in Bauwens and Lubrano (1998). In the graphs, a $\pm 10\%$ confidence band is displayed as explained in Bauwens and Lubrano (1998).

Convergence results were very good for 1979 and 1988, good for 1996. But, there were some convergence problems for 1992. So we had for that case to increase the warming up of the chain from 1000 to 10 000 draws. Posterior moments are presented in Table 4 with posterior standard deviation in parenthesis.

We can identify three social classes characterized by their posterior mean income (computed by averaging the transformation of the draws $\exp(\mu_i^{(l)} + \sigma^{2(l)}/2)$) given in Table 5. The first group of the mixture, those with the lowest mean income represents 30% of the sample on average. The second group represents the highest proportion of the population, 68% on average and has a much higher income, representing roughly twice of that of the poorest group. The last group, having the highest mean income, is much smaller, 2% on average of the whole sample. These posterior characteristics give arguments to interpret our mixture model as representing a heterogeneous population where each group corresponds to a particular income group. Of course, this is only an interpretation. The mean income of the first group increased very slowly over 1979–1992 and had a sharp increase in 1996. The income rise was more regular for the second group while the third group experienced a huge increase in income over 1979–1992, with a sharp decrease in 1996.

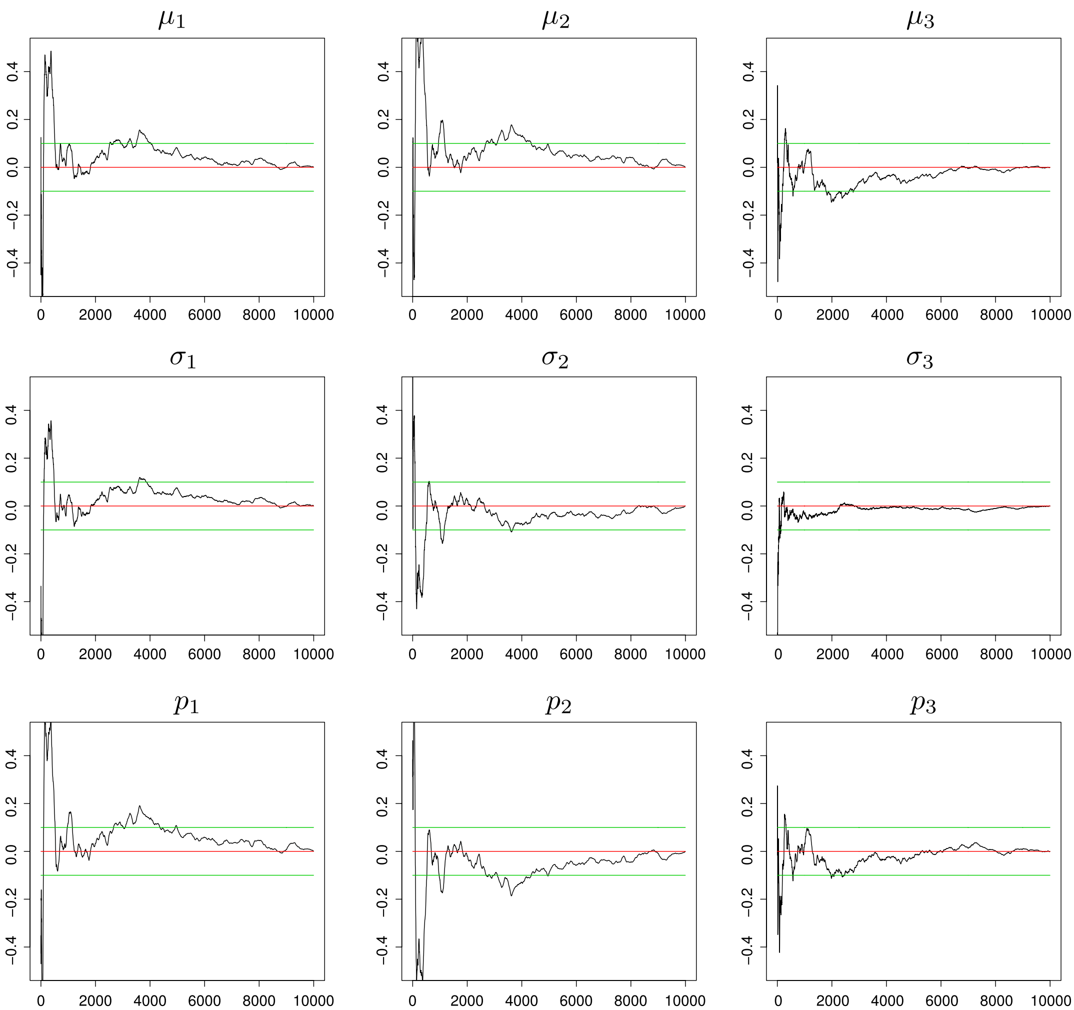

Fig. 1. Cumsum graphs for 1979 using an asymmetric informative prior.

Table 5
Posterior mean income per group.

	Group 1	Group 2	Group 3
1979	50.85	96.54	101.79
1988	55.82	120.02	187.51
1992	58.10	129.35	264.70
1996	70.90	132.36	208.50

The probability weights of the mixture can be used as a tool to provide information on the evolution of the income mobility between each group, provided each member of the mixture still represents the same group of persons. If we stick to that interpretation (which could be guaranteed only if we had real panel data) Table 4 reveals an important change in structure for the three identified groups. The importance of the low income group remained constant over 1979–1992, but increased slightly in 1996. The high income group was the most affected by the fight against inequality that followed the Thatcher period. Its importance increased in 1988–1992, but returned to its initial level in 1996. The importance of the middle group decreased steadily over the period. These findings are in lines with Jenkins (1996) who claimed that, during the late 1970s, "*the distinct clump in the concentration of people around middle income levels began to break up and polarize towards high and low incomes*", giving support to opposed opinions and claims. Mr. Kinnock (British Labour Party) argued that "*While the very rich have lost some of their riches to the less rich, over time, the poor have hardly profited proportionately*" (The Future of Socialism, Fabian Tract No. 506, January 1986). On the other side, the view of Mrs. Thatcher claimed that "*the real incomes have increased throughout all income groups*" (Weekly Hansard, 27 April 1989). From Table 5, we note that the greatest increase is for the upper class.

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

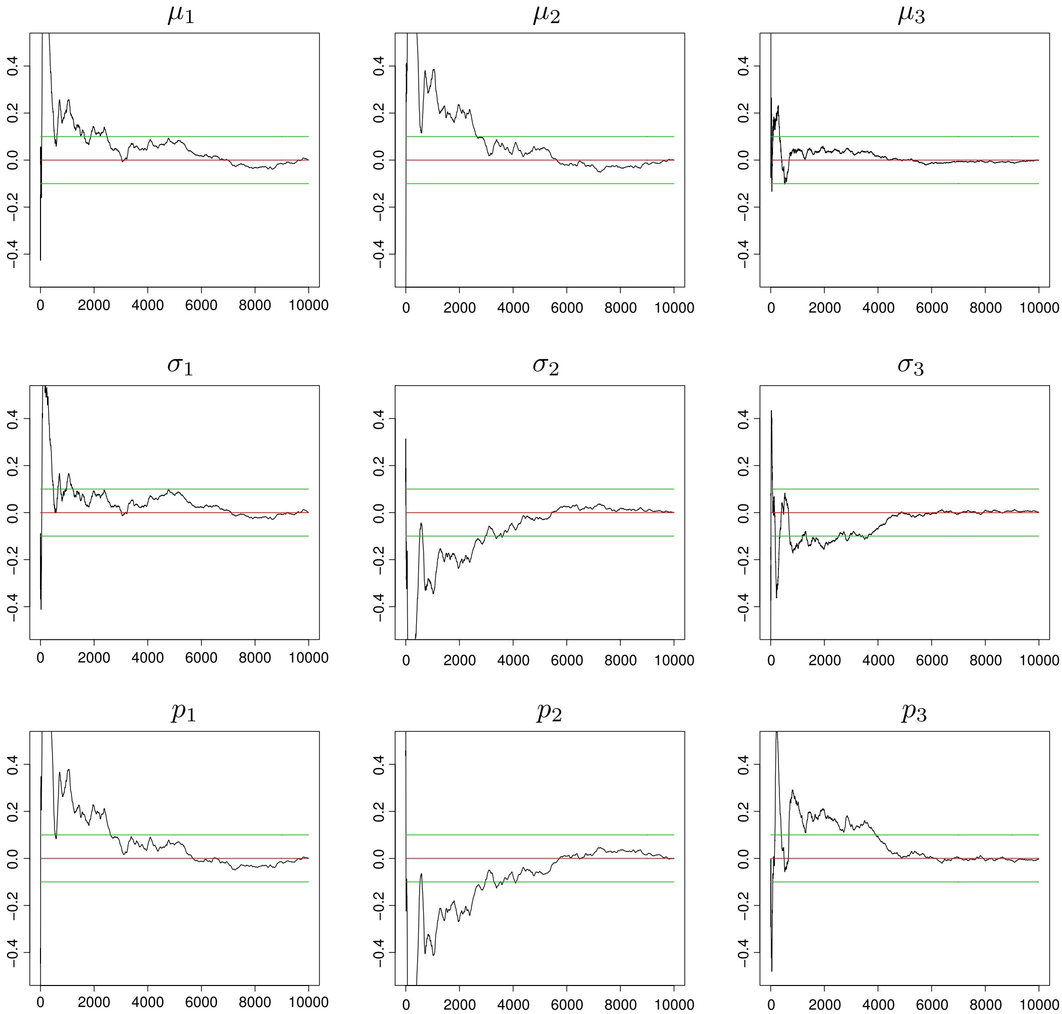

Fig. 2. Cumsum graphs for 1988 using an asymmetric informative prior.

Table 6
Posterior means and standard deviations of inequality indices.

	GE $\alpha = 0.5$	Theil	MLD	Atkinson $\epsilon = 0.5$	Gini
1979	0.109 (0.0034)	0.113 (0.0053)	0.110 (0.0027)	0.056 (0.0031)	0.259 (0.0022)
1988	0.164 (0.0073)	0.179 (0.0120)	0.161 (0.0055)	0.085 (0.0050)	0.309 (0.0032)
1992	0.204 (0.0171)	0.246 (0.0363)	0.196 (0.0115)	0.108 (0.0109)	0.320 (0.0031)
1996	0.154 (0.0096)	0.170 (0.0190)	0.152 (0.0066)	0.079 (0.0066)	0.298 (0.0029)

As the Atkinson index is only decomposable in an indirect way, we have used the same calculation as for the Gini index, using the general expression of the Atkinson index given in Table 1.

4.4. Inequality growth in the UK from 1979 to 1996

Let us now summarize the overall densities and their decomposition into sub-groups by means of inequality indices. Table 6 gives the posterior mean for the Generalized Entropy index (GE) for $\alpha = 0.5$, the Theil index, the Mean Logarithmic Deviation (MLD) index, the Atkinson index for $\epsilon = 0.5$ and the Gini index, with their posterior standard deviations between brackets. There is a considerable increase in all the inequality indices from 1979 to 1992, a fact that should be related to the period 1979–1990 when Margaret Thatcher was Prime Minister. This inequality growth is slowing down between 1988 and 1992. And from 1992 to 1996, all these inequality measures decrease, going back to their levels of somewhere between 1988 and 1992. These results are in lines with the annual report of the Department of Social Security (1998), see Jenkins (1996, 2000).

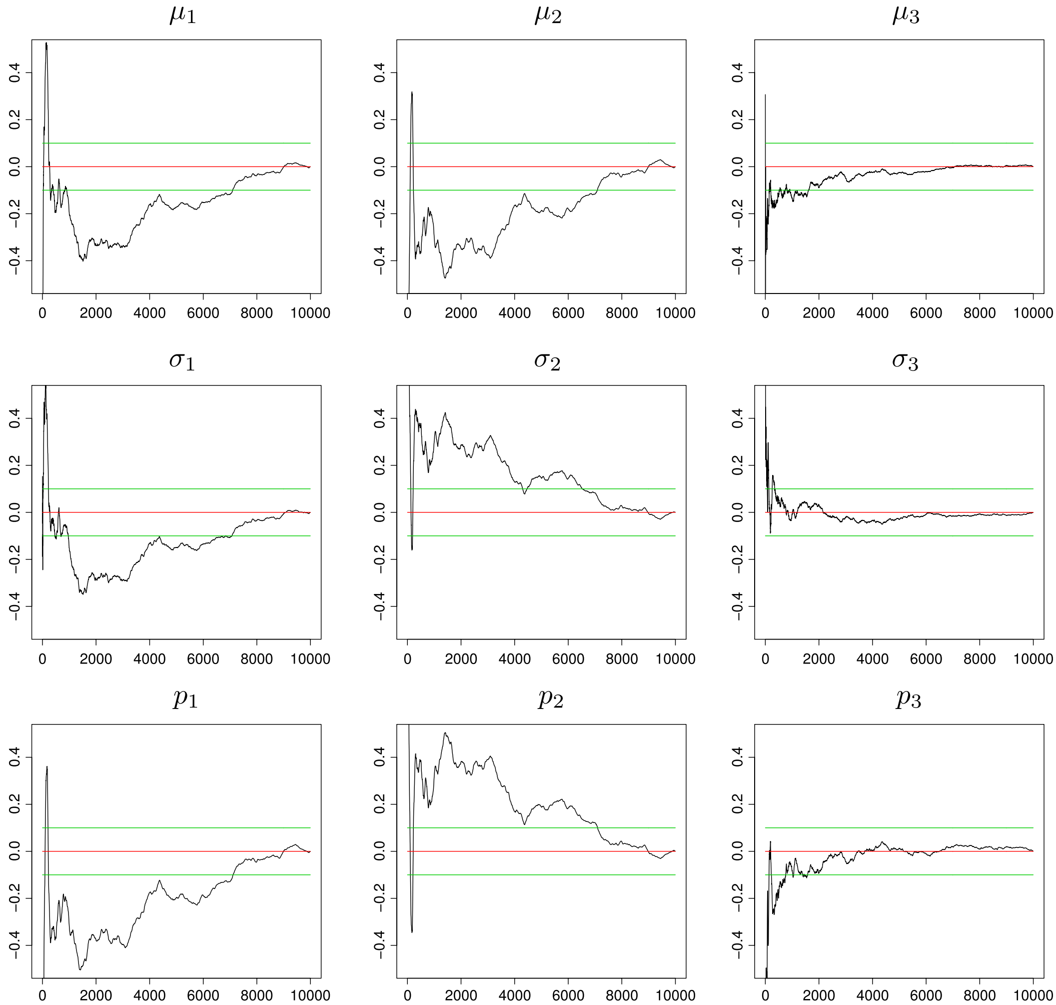

Fig. 3. Cumsum graphs for 1992 using an asymmetric informative prior.

Table 7
MLD decomposition of the income inequality.

Year	Within		Between		Total
	Estimates	Proportion	Estimates	Proportion	Estimates
1979	0.071 (0.0033)	0.65	0.039 (0.0032)	0.35	0.110 (0.0027)
1988	0.102 (0.0046)	0.64	0.059 (0.0051)	0.36	0.161 (0.0055)
1992	0.124 (0.0068)	0.64	0.072 (0.0096)	0.36	0.196 (0.0115)
1996	0.108 (0.0058)	0.71	0.044 (0.0068)	0.29	0.152 (0.0066)

Let us now decompose the MLD inequality measure among the three groups that have been identified. The within-subgroups inequality represents on average 65% of total inequality and this proportion does not vary much over time, except for the last period 1996. The within-subgroup inequality increased from 1979 to 1992 and then decreased, while the between-subgroup inequality increased sharply from 1979 to 1992 and decreased sharply after that period to reach a level comparable to its initial value (see Table 7).

In Table 8, we give the evolution of inequality within the three income groups. Within the group of lower income households, inequality is low, but increases steadily. Within the middle income group, inequality is slightly larger, but follows the same pattern. Finally, inequality within the upper income group is ten times that of the intermediate group. It increases over 1979–1992, but decreases in 1996.

We should however be cautious about this interpretation concerning the evolution of inequality within and between groups. We have interpreted each member of the mixture as representing a social group. This is already a strong interpretation. And we have no guarantee that the mixture members identify exactly the same group of people in each period. In order to fully describe social mobility, we would need panel data and dynamic models.

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

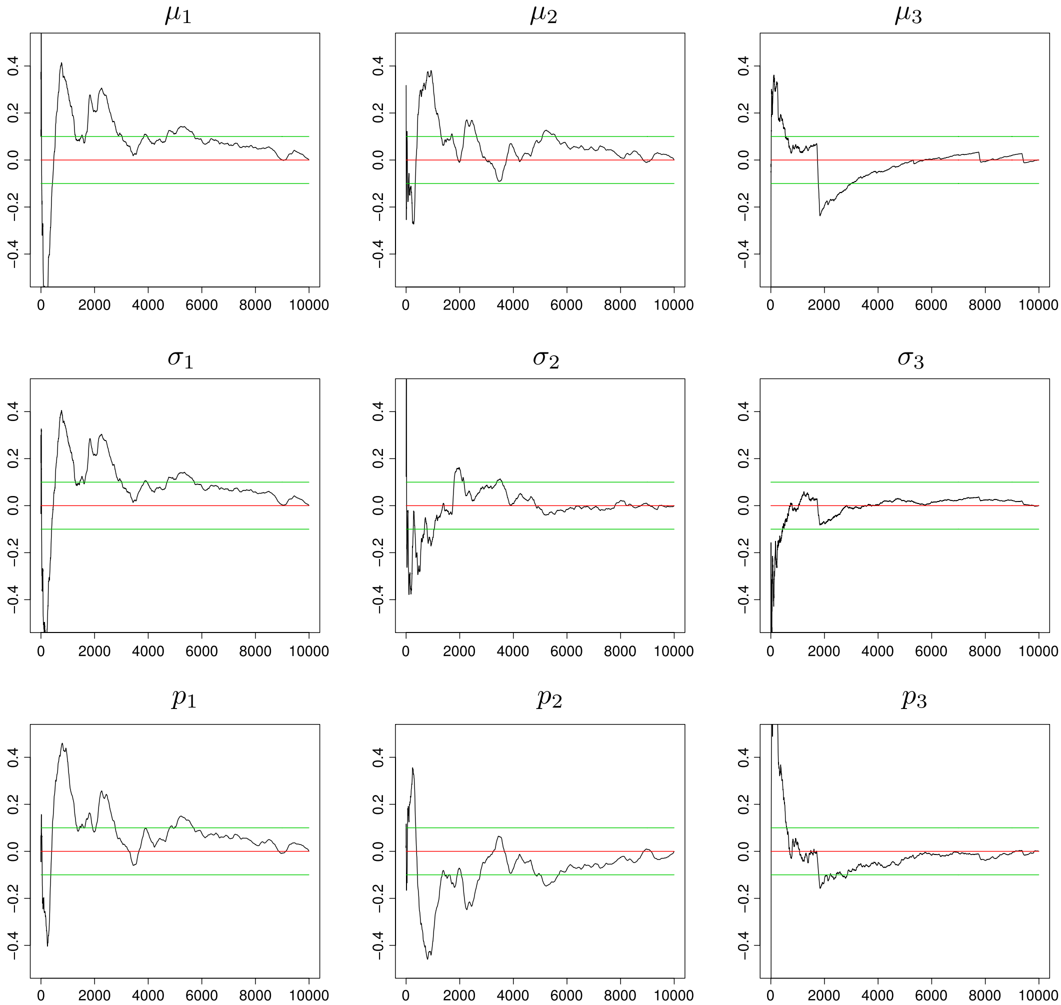

Fig. 4. Cumsum graphs for 1996 using an asymmetric informative prior.

Table 8
MLD within-group inequality.

	Group 1	Group 2	Group 3
1979	0.0310 (0.0023)	0.0779 (0.0036)	0.7438 (0.1224)
1988	0.0339 (0.0026)	0.1040 (0.0051)	0.7602 (0.1085)
1992	0.0421 (0.0033)	0.1114 (0.0066)	1.4374 (0.1835)
1996	0.0543 (0.0054)	0.1160 (0.0071)	1.2353 (0.2112)

4.5. Comparing classical and Bayesian standard deviations

Several methods were proposed in the classical literature to compute the standard deviation of a Gini index. The question is complex because the Gini index is based on ordered data. Building on the fact that a Gini index can be seen as a covariance between observations and their rank, Giles (2004) proposes computing the Gini index using a linear regression, which is correct, and to compute its associated standard error using the standard error of the regression, which appears to be misleading because the usual assumptions underlying the OLS are not satisfied. Davidson (2009) proposed an asymptotic method based on the natural estimator of the cumulative distribution. We are now in a position, for the particular samples of the FES, to compare these methods with our Bayesian approach. According to our computations reported in Table 9, it appears that all methods give comparable results for the value of the index, but that they differ in their standard deviations. Presumably because of the strong prior information we introduced, our Bayesian standard deviations for the Gini index are in general slightly lower than their asymptotic counterpart. The regression method, on the other side, gives much higher classical standard errors. It is well known that Giles' regression corrects for heteroscedasticity, but not for autocorrelation, which leads to biased standard errors.

Please cite this article in press as: Lubrano, M., Ndoye, A.A.J., Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.10.009

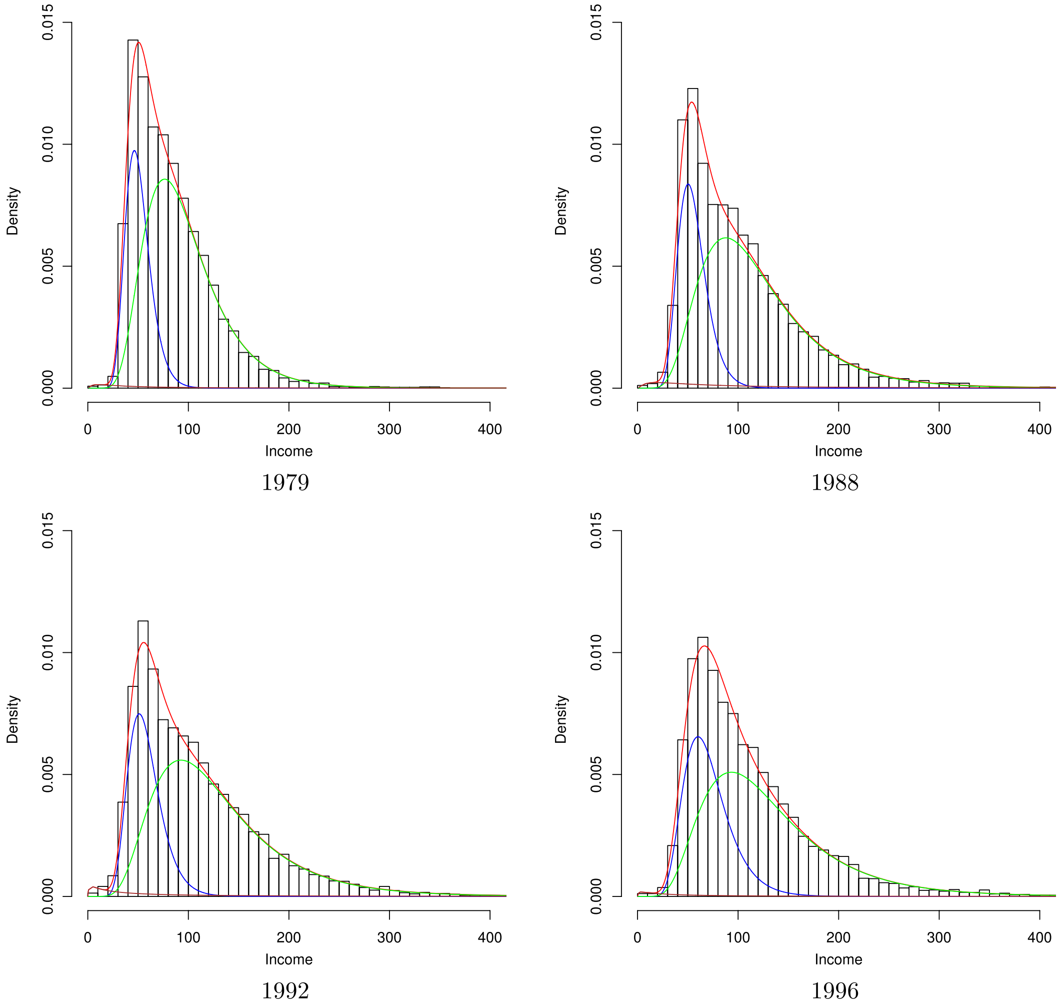

Fig. 5. MCMC predictive densities for the FES data.

Table 9
Different methods for estimating standard deviations of a Gini index.

	Bayesian	Asymptotic, Davidson (2009)	Regression, Giles (2004)
1979	0.259 (0.0022)	0.256 (0.0023)	0.256 (0.0058)
1988	0.309 (0.0032)	0.307 (0.0034)	0.307 (0.0053)
1992	0.320 (0.0031)	0.322 (0.0037)	0.322 (0.0053)
1996	0.298 (0.0029)	0.297 (0.0032)	0.297 (0.0058)

5. Conclusion

A mixture of log-normal distributions was found to be a convenient and powerful explanatory model of the UK income distribution. We have demonstrated how a Gibbs sampler can be used to estimate this type of mixture when we elicit more precise prior information which helps to reduce the usual label switching problem. Using the UK FES data, we have managed to identify and characterize income groups.

We were able, in this context, to provide a Bayesian inference for commonly used inequality indices that are decomposable. Using a Rao-Blackwellization, we could provide a plausibly more numerically accurate evaluation of posterior standard deviations. We have extended the method to indices that are not decomposable at the price of a one dimensional numerical integration, showing how it works for the Gini and the Atkinson indices.

As a final remark, we can note by inspecting the graph of the posterior predictive density (see Fig. 5) that the last member of the log-normal mixture does not seem satisfactory for modelling high incomes. In order to have a large right tail for the third group, we must have a large value for σ_3^2. A Pareto tail would have been intellectually more satisfactory, corresponding to a hybrid mixture of two log-normals and a Pareto. Hybrid mixtures are not common in the literature. This topic is left for future research.

Acknowledgements

We are grateful to Emmanuel Flachaire for his detailed comments and for sharing data. Our thanks go to Karim Abadir, Stephen Bazen, Christian Robert for helpful comments and suggestions. We thank also the participants of the Rimini Bayesian Workshop 2011, the participants of the European Seminar on Bayesian Econometrics 2011 held in Brussels, in particular Sylvia Frühwirth-Schnatter and the participants of the Third French Econometrics Conference 2011 held in Aix-en-Provence for comments. We particularly appreciated the two referees for their extended comments and suggestions. Of course, the usual disclaimers apply. The help of Stephen Bazen was particularly helpful for proof reading. The revision of this paper benefited from the support of the A*MIDEX project (no ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR).

References

Abadir, K.M., 1999. An introduction to hypergeometric functions for economists. Econometric Rev. 18 (3), 287–330.
Aitchison, J., Brown, J.A.C., 1957. The Lognormal Distribution. Cambridge University Press, London.
Bahadur, R.R., Savage, L.J., 1956. The nonexistence of certain statistical procedures in nonparametric problems. Ann. Statist. 27, 1115–1122.
Bauwens, L., Lubrano, M., 1998. Bayesian inference on GARCH models using the Gibbs sampler. Econom. J. 1 (1), 23–46.
Bauwens, L., Lubrano, M., Richard, J., 1999. Bayesian Inference in Dynamic Econometric Models. Oxford University Press.
Biewen, M., 2002. Bootstrap inference for inequality, mobility and poverty measurement. J. Econometrics 108 (2), 317–342.
Biewen, M., Jenkins, S.P., 2006. Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case. Oxford Bull. Econ. Stat. 68 (11), 371–383.
Bourguignon, F., 1979. Decomposable income inequality measures. Econometrica 47 (4), 901–920.
Bourguignon, F., 2012. La Mondialisation de l'Inégalité. Seuil, Paris.
Celeux, G., Forbes, F., Robert, C., Titterington, D., 2006. Deviance information criteria for missing data models. Bayesian Anal. 1 (4), 651–674.
Celeux, G., Hurn, M., Robert, C.P., 2000. Computational and inferential difficulties with mixture posterior distributions. J. Amer. Statist. Assoc. 95 (451), 957–970.
Chib, S., 1995. Marginal likelihood from the Gibbs output. J. Amer. Statist. Assoc. 90 (432), 1313–1321.
Chotikapanich, D., Griffiths, W.E., 2008. Estimating income distributions using a mixture of gamma densities. In: Chotikapanich, D. (Ed.), Modeling Income Distributions and Lorenz Curves. Springer Verlag, Heidelberg, pp. 285–302. (Chapter 16).
Cowell, F., 1995. Measuring Inequality. Prentice Hall, London.
Davidson, R., 2009. Reliable inference for the Gini index. J. Econometrics 150 (1), 30–40.
Davidson, R., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68 (6), 1435–1464.
Davidson, R., Flachaire, E., 2007. Asymptotic and bootstrap inference for inequality and poverty measures. J. Econometrics 141 (1), 141–166.
Diebolt, J., Robert, C.P., 1994. Estimation of finite mixture distributions through Bayesian sampling. J. Roy. Statist. Soc. Ser. B 56 (2), 363–375.
Ferguson, T.S., 1983. Bayesian density estimation by mixtures of normal distributions. In: Rizvi, H., Rustagi, J. (Eds.), Recent Advances in Statistics. Academic Press, New York.
Flachaire, E., Nunez, O., 2002. Estimation of the income distribution and detection of subpopulations: an explanatory model. Comput. Statist. Data Anal. 51 (7), 3368–3380.
Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. J. Amer. Statist. Assoc. 96 (453), 194–209.
Frühwirth-Schnatter, S., 2004. Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. Econom. J. 7 (1), 143–167.
Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. In: Springer Series in Statistics, Springer, Heidelberg.
Giles, D.E.A., 2004. Calculating a standard error for the Gini coefficient: some further results. Oxford Bull. Econ. Stat. 66, 425–433.
Green, P., 1995. Reversible jump MCMC computation and Bayesian model determination. Biometrika 82 (4), 711–732.
Jasra, A., Holmes, C.C., Stephens, D.A., 2005. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. Statist. Sci. 20 (1), 50–67.
Jenkins, S.P., 1996. Recent trends in the UK income distribution: what happened and why? Oxford Rev. Econ. Policy 12 (1), 29–46.
Jenkins, S.P., 2000. Trends in the UK income distribution. In: Hauser, R., Becker, I. (Eds.), The Personal Distribution of Income in an International Perspective. Springer-Verlag, Berlin.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90 (430), 773–795.
Marin, J.-M., Mengersen, K., Robert, C.P., 2005. Bayesian modelling and inference on mixtures of distributions. In: Dey, D., Rao, C. (Eds.), Bayesian Thinking Modeling and Computation. In: Handbook of Statistics, vol. 25. Elsevier, pp. 459–507.
Marin, J., Robert, C., 2007. Bayesian Core: A Practical Approach to Computational Bayesian Statistics. Springer Texts in Statistics. Springer Verlag, Heidelberg.
Marron, J.S., Schmitz, H.P., 1992. Simultaneous density estimation of several income distributions. Econometric Theory 8 (4), 476–488.
McDonald, J.B., 1984. Some generalized functions for the size distribution income. Econometrica 52 (3), 647–663.
McDonald, J., Ransom, M., 1979. Functional forms, estimation techniques and the distribution of income. Econometrica 47 (6), 1513–1525.
Mitzenmacher, M., 2004. A brief history of generative models for power law and lognormal distributions. Internet Math. 1 (2), 226–251.
Paap, R., van Dijk, H.K., 1998. Distribution and mobility of wealth of nations. Eur. Econ. Rev. 42, 1269–1293.
Reed, W.J., Jorgenson, M., 2004. The double Pareto-lognormal distribution—a new parametric model for size distributions. Comm. Statist. Theory Methods 33 (8), 1733–1753.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.
Shorrocks, A.F., 1980. The class of additively decomposable inequality measures. Econometrica 48 (3), 613–625.
Singh, S.K., Maddala, G.S., 1976. A function for size distribution of income. Econometrica 44 (5), 963–970.
Spiegelhalter, D.J., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B 64 (4), 583–639.
Stephens, M., 2000a. Bayesian analysis of mixture models with an unknown number of components: an alternative to reversible jump methods. Ann. Statist. 28, 40–74.
Stephens, M., 2000b. Dealing with label switching in mixture models. J. R. Stat. Soc. Ser. B 62 (4), 795–809.
Yu, B., Mykland, P., 1998. Looking at Markov samplers through CUMSUM path plots: a simple diagnostic idea. Stat. Comput. 8, 275–286.