Bayesian Assessment of Lorenz and Stochastic Dominance
using a Mixture of Gamma Densities

David Lander
*University of Melbourne*

William Griffiths
*University of Melbourne*

Duangkamon Chotikapanich
*Monash University*

August 26, 2012

Preliminary: Please do not quote without authors’ permission
Abstract

Hypothesis tests for dominance in income distributions have received considerable attention in recent literature. See, for example, McCaig and Yatchew (2007), Barrett and Donald (2003), Davidson and Duclos (2000) and references therein. Such tests are useful for assessing progress towards eliminating poverty and for evaluating the effectiveness of various policy initiatives directed towards welfare improvement. To date the focus in the literature has been on sampling theory tests. In such tests we either reject a null hypothesis, implying dominance does exist, or we conclude there is insufficient evidence to reject a null hypothesis, implying the evidence is inadequate to prove dominance. Chotikapanich and Griffiths (2006) develop and apply Bayesian methods of inference to problems of Lorenz and stochastic dominance. In this paper we extend their method, applying it to a more flexible class of income distributions. Outcomes from income distribution comparisons are reported in terms of the posterior probability that dominance exists. Reporting results about uncertain outcomes in terms of probabilities has the advantage of being more informative than a simple reject / do-not-reject outcome. Whether a probability is sufficiently high or low for a policy maker to take a particular action is a decision for that policy maker.

The methodology is applied to data for Indonesia from the Household Expenditure Survey for the years 1999, 2002, 2005 and 2008. We assess the likelihood of dominance from one time period to the next. In this paper income is assumed to follow a mixture of gamma distributions. The posterior probability of dominance is given by the proportion of times one distribution dominates the other over the posterior observations generated by Markov chain Monte Carlo.
1. Introduction

Statistical tests of dominance regularly appear in the economics and finance literature. These tests employ statistical methods to compare the distributions of two random variables (or one random variable at two points in time) in such a way as to determine if one ‘dominates’ another. These tests have been used to compare income distributions in welfare analysis, the distribution of asset returns in portfolio analysis, and risk analysis in actuarial science. This paper will focus on income distributions and welfare analysis. For applications of stochastic dominance to asset returns in portfolio analysis see Wong et al. (2008) and Sriboonchitta et al. (2010); and for actuarial sciences and risk analysis see Kaas et al. (1994) and Denuit et al. (2005).

Due to the diverse and useful applications for these tests, their degree of sophistication has improved considerably in the past few decades. This paper contributes to the empirical stochastic dominance literature by developing a Bayesian approach to testing, rather than continuing the current trend of further developing the sampling-theory approach.

Atkinson (1970) and Shorrocks (1983) laid the theoretical foundations for using stochastic dominance-based approaches to analyse social welfare. See Lambert (1993) for a useful summary of the history of stochastic dominance and the theory of income distributions. Among other things, they showed that Lorenz dominance can be interpreted in terms of social welfare for a general class of income-utility functions. Specifically this relationship holds for monotonically increasing concave, but otherwise arbitrary, income-utility functions. This was an important result, since it implied that the stochastic dominance criterion is robust to changes in the functional form of the social welfare or poverty index. Allowing practitioners to make strong conclusions and avoid specifying the functional form of the utility or poverty function is a compelling reason to use stochastic dominance conditions over other inequality measures. However, other inequality measures, such as the Gini coefficient and Atkinsons inequality index are still used when income distributions cannot be ordered according to stochastic dominance criteria. Of course this involves placing more restrictive assumptions on the functional form of the social welfare or poverty index. For details of these various concepts, and the relationships between them, see Atkinson and Bourguignon (1982), Lambert (1993), and Maasoumi (1997).

The empirical study of stochastic dominance is typically undertaken in a sampling-theory framework using hypothesis tests; this involves using a modified Kolmogorov-Smirnov (KS)
statistic, or by undertaking simultaneous tests at multiple points. The KS statistic-based tests have been proposed and developed by McFadden (1989), Klecan et al. (1991), Kaur et al. (1994), Maasoumi and Heshmati (2000), Barrett and Donald (2003), and Linton et al. (2005). The main issue associated with these tests is the determination of critical values. The other type of tests include those suggested by Anderson (1996) and Davidson and Duclos (2000). These tests may be inconsistent, but they do not require bootstrapping, simulation, or subsampling techniques to determine critical values. These tests have also been modified for multivariate analysis by Crawford (2005) and Duclos et al. (2004) to make multiple comparisons on a grid of points. Crawford extends Anderson (1996), while Duclos et al. extends Davidson and Duclos (2000). These procedures involve calculating a t-statistic at each point of comparison and comparing it with the appropriate critical value obtained from the studentized maximum modulus (SMM). Along with the usual issue of inconsistency, Tse and Zhang (2004) identified that the nominal size of the SMM test overstates the actual size when the tests are not independent. More recently, McCaig and Yatchew (2007) analyzed cross-country welfare using multidimensional stochastic dominance hypotheses; Lean et al. (2008) examined the size and power of several stochastic dominance tests, concluding that the test developed by Davidson and Duclos has better size and power performance than the tests developed by Anderson and Kaur et al.; and Berrendero and Carcamo (2011) used convex and concave-type orders to derive stochastic dominance test statistics as functions of L-statistics.

This paper defines a Bayesian approach to dominance tests, and applies this method using income distribution data. This approach is novel since it deviates from the standard sampling-theory tests. The advantage of the Bayesian approach over the sampling-theory approach is that it does not require specification of a null and alternative hypothesis, test statistic and level of significance. This avoids the potential problems of subjectivity influencing analysis through the specification of these things, and the generally unavoidable problem of hypothesis test results not being invariant to re-specification of the null and alternative hypotheses. The main disadvantage of the Bayesian approach is that the posterior probabilities for dominance will depend on how the income distribution is modelled through the likelihood function, and the prior information that is placed on unknown parameters. The issue of subjectivity relating to the specification of prior information, however, can be limited by using relatively uninformative priors. Furthermore, since the weight placed on the prior information is reduced when the size of the dataset is increased,
the priors are unlikely to be a significant determinant of the posterior probabilities. The choice of likelihood function, however, is very important. For example, Chotikapanich and Griffiths (2006) found that their results were quite sensitive to the chosen distribution. They employed two commonly-used distributions for assessing income data as their likelihood function – the Dagum and Singh-Maddala distributions – and concluded that a relatively flexible likelihood function is necessary for robust results.

This paper extends the analysis of Chotikapanich and Griffiths by choosing a more flexible likelihood function: a mixture of gamma distributions. A gamma mixture is specifically chosen to analyze data defined over the positive reals \((0, \infty)\) – an appropriate choice for income distributions. The added flexibility achieved by a mixture, however, makes the dominance tests significantly more complex. We are confronted with the problem of choosing an arbitrary number of points to assess dominance across the \((0, \infty)\) interval, or having to employ numerical methods to invert the distribution function of the mixture, which enables us to use the quantile function and restrict our analysis to the \((0, 1)\) interval. For a similarly sized grid of points, numerically deriving the quantile function of the mixture is significantly more computationally demanding than using the distribution function approach.

The remainder of the paper is structured as follows. We begin by specifying the methods of dominance that will be considered in Section 2. A gamma mixture model and its associated Markov chain Monte Carlo (MCMC) algorithm for drawing observations from the joint posterior density for the parameters of the mixture is defined in Section 3. Data and estimation results for some Indonesian income distributions are considered in Section 4, with dominance results presented in Section 5. Section 6 offers concluding remarks.

2. Dominance Conditions and Bayesian Assessment
To introduce Lorenz, generalized Lorenz and first order stochastic dominance consider an income distribution that is described by density and distribution functions \(f_X(x)\) and \(F_X(x)\), respectively. Also, assume that mean income \(\mu_X = E(X)\) is finite. The Lorenz curve that gives the proportion of total income earned by the poorest proportion \(u\) of the population is given by
We say that an income distribution for \( X \) Lorenz dominates (LD) a distribution for \( Y \) (say), \( X \geq_{LD} Y \), if and only if
\[
L_X(u) \geq L_Y(u) \quad \text{for all } 0 \leq u \leq 1
\] (2)

While this definition is the typical one used in the economics literature (see, for example, Lambert (1993) and Barrett and Donald (2003b)), the definition used in much of the statistics literature follows the opposite convention, with \( L_Y(u) \geq L_X(u) \) being the condition for \( X \geq_{LD} Y \). See, for example, Kleiber and Kotz (2003). Since \( L_X(u) \geq L_Y(u) \) implies higher welfare for distribution \( X \) in the sense that, other thing equal, less inequality is preferred to more inequality, we refer to this condition as one where \( X \) dominates \( Y \).

Because Lorenz dominance considers only the degree of inequality and not the level of income, and higher levels of income are associated with higher levels of welfare, another dominance relation known as generalized Lorenz dominance (GLD) is often considered. We say that \( X \) generalized-Lorenz dominates \( Y \), written as \( X \geq_{GLD} Y \) if and only if
\[
\mu_X L_X(u) \geq \mu_Y L_Y(u) \quad \text{for all } 0 \leq u \leq 1
\] (3)

Given the expression for the Lorenz curve in equation (1), the condition in (3) can also be expressed as
\[
\int_0^u t^{-1} \int_0^u F_X(x) dt \geq \int_0^u t^{-1} \int_0^u F_Y(x) dt \quad \text{for all } 0 \leq u \leq 1
\] (4)

Writing the relation for generalized Lorenz dominance (GLD) in this way demonstrates its equivalence to second order stochastic dominance (SSD). See, for example, Maasoumi (1997) or Kleiber and Kotz (2003, p.25). An equivalent condition for SSD, and one that turns out to be convenient in our later analysis is
\[
\int_0^x t^{-1} \int_0^x F_X(x) dt \leq \int_0^x t^{-1} \int_0^x F_Y(x) dt \quad \text{for all } 0 < x < \infty
\] (5)

A stronger condition for welfare improvement than SSD (GLD) is that of first-order stochastic dominance (FSD). The distribution for \( X \) first-order stochastically dominates \( Y \), written
\[ X \geq_{\text{FSD}} Y, \text{ if and only if} \]

\[ F_X^{-1}(u) \geq F_Y^{-1}(u) \quad \text{for all } 0 \leq u \leq 1 \quad (6) \]

In this case the level of income from distribution \( X \) is greater than the level of income from distribution \( Y \) for all population proportions \( u \). Alternatively, \( X \geq_{\text{FSD}} Y, \text{ if and only if} \]

\[ F_X(x) \leq F_Y(x) \quad \text{for all } 0 < x < \infty \quad (7) \]

Given two distributions, each with known parameter values, one way to assess each form of dominance (LD, GLD or FSD) is to compute \( L(u), \mu L(u) \) and \( F^{-1}(u) \) for both distributions for a grid of values for \( u \) in the interval \((0,1)\). If the grid contains a relatively large number of values, and the dominance inequality being considered is satisfied for all those values, then it is reasonable to conclude that the condition is satisfied for all \( u \), and hence dominance holds. Alternatively, for GLD and FSD one could compute \( F(x) \) and \( \int_0^x F(t)dt \) for both distributions for a grid of values of \( x \) in the interval \((0,x_{\text{max}})\) where \( x_{\text{max}} \) is a value deemed to be sufficiently large, and examine whether the relevant inequalities are satisfied for all these values.

Suppose now that the distributions for \( X \) and \( Y \) have unknown parameter vectors \( \theta_X \) and \( \theta_Y \) that are estimated using income distribution data. Since these parameters are not known with certainty, any conclusion about whether one distribution dominates another cannot be made with certainty. In Bayesian inference uncertainty about whether one distribution dominates another can be expressed in terms of a probability statement. To obtain (estimate) such a probability statement, assume we have draws on \( \theta_X \) and \( \theta_Y \) from the two posterior densities \( p(\theta_X | x) \) and \( p(\theta_Y | y) \). Examining FSD as an example, the probability that \( F_X(x;\theta_X) \leq F_Y(x;\theta_Y) \) for a given value \( x \) is given by the proportion of values of \( \theta_X \) and \( \theta_Y \) for which the inequality holds.

The probability \( X \geq_{\text{FSD}} Y \) within a range \((0,x_{\text{max}})\) is given by the proportion of values \( \theta_X \) and \( \theta_Y \) for which the inequality holds for all values of \( x \). In practice we can consider a grid of \( x \) values within the interval and count the number of parameter draws where the inequality holds for all \( x \) in the grid. A finer grid can be taken in the region that counts: those values of \( x \) where the probability of \( F_X(x;\theta_X) \leq F_Y(x;\theta_Y) \) reaches a minimum. Similar probability statements can
be made for LD and GLD.

3. Bayesian Estimation of a Gamma Mixture

Since the probability of dominance could depend heavily on the assumed functional forms for the income distributions, it is important to specify functional forms that are sufficiently flexible to fit well. Mixtures of densities generally have the ability to approximate any functional form. We chose a gamma mixture. The more popular lognormal was not used since its predictive density does not have moments if a noninformative prior is used. An income distribution that follows a gamma mixture with \(k\) components can be written as

\[
 f(x \mid \mathbf{w}, \mathbf{\mu}, \mathbf{v}) = \sum_{z=1}^{k} w_z G(x \mid v_z, \mu_z) \tag{8}
\]

where \(x\) is a random draw of income from the probability density function (pdf) \(f(x \mid \mathbf{w}, \mathbf{\mu}, \mathbf{v})\), with parameter vectors, \(\mathbf{w} = (w_1, w_2, \ldots, w_k)'\), \(\mathbf{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)'\), and \(\mathbf{v} = (v_1, v_2, \ldots, v_k)'\). The pdf \(G(x \mid v_z, \mu_z)\) is a gamma density with mean \(\mu_z > 0\) and shape parameter \(v_z > 0\). That is,

\[
 G(x \mid v_z, \mu_z) = \frac{(v_z/\mu_z)^{v_z}}{\Gamma(v_z)} x^{v_z-1} \exp\left(-\frac{v_z}{\mu_z} x\right) \tag{9}
\]

Including the mean \(\mu_z\) as one of the parameters in the pdf makes the parameterization in (9) different from the standard textbook one, but it is convenient for later analysis. The parameter \(w_z\) is the probability that the \(i\)-th observation comes from the \(z\)-th component in the mixture. To define it explicitly, let \(x = (x_1, x_2, \ldots, x_n)\) be a random sample from (8), and let \(Z_1, Z_2, \ldots, Z_n\) be indicator variables such that \(Z_i = z\) when the \(i\)-th observation comes from the \(z\)-th component in the mixture. Then,

\[
 P(Z_i = z \mid \mathbf{w}) = w_z \quad \text{for } z = 1, 2, \ldots, k
\]

with \(w_z > 0\) and \(\sum_{z=1}^{k} w_z = 1\). Also, conditional on \(Z_i = z\), the distribution of \(x_i\) is \(G(v_z, v_z/\mu_z)\).

To use Bayesian inference, we specify prior distributions on the unknown parameters \(\mathbf{w}\), \(\mathbf{\mu}\), and \(\mathbf{v}\), and then combine these pdfs with the likelihood function defined by (8) to obtain a joint posterior pdf for the unknown parameters. This joint posterior pdf represents our post-
sample knowledge about the parameters and is the source of inferences about them. However, as is typically the case in Bayesian inference, the joint posterior pdf is analytically intractable. This problem is solved by using MCMC techniques to draw observations from the joint posterior pdf and using these draws to estimate the quantities required for inference. Because we are interested in not just the parameters, but also the income distribution, the Gini coefficient, and the Lorenz curve, the parameter draws are also used in further analysis to estimate posterior information about these quantities.

The MCMC algorithm used to draw observations from the posterior density for \((\mu, v, w)\) is taken from Wiper, Rios Insua and Ruggeri (2001). In the context of other problems, Wiper et al. consider estimation for both a known and an unknown \(k\). We will assume a known value of \(k\) that is specified \textit{a priori}. In our empirical work we considered values of \(k\) up to 5 and settled on \(k = 5\) as an adequate formulation. The MCMC algorithm is a Gibbs’ sampling one where draws are taken sequentially and iteratively from the conditional posterior pdfs for each of the parameters. Because only the conditional posterior pdfs are involved in this process, it is not necessary to specify the complete joint posterior pdf. The relevant conditional posterior pdfs are sufficient; they are specified below after we introduce the prior pdfs.

Following Wiper et al. (2001), the prior distributions used for each of the parameters are

\[
\begin{align*}
  f(w) &= D(\gamma) \propto w_1^{\phi_1-1} w_2^{\phi_2-1} \cdots w_k^{\phi_k-1} \\
  f(v_z) &= \exp\{-\theta v_z\} \quad \text{(exponential)} \\
  f(\mu_z) &= GI(\alpha_z, \beta_z) \propto \mu_z^{(\alpha_z-1)} \exp\left\{ -\frac{\beta_z}{\mu_z} \right\} \quad \text{(inverted gamma)}
\end{align*}
\]

for \(z = 1, 2, \ldots, k\)

The Dirichlet distribution is the same as a beta distribution for \(k = 2\) and a multivariate extension of the beta distribution for \(k > 2\). Its parameters are \(\phi = (\phi_1, \phi_2, \ldots, \phi_k)^T\). To appreciate the relationship between the gamma and inverted gamma pdfs, note that if \(y \sim G(\alpha, \beta)\), then \(q = (1/y) \sim GI(\alpha, \beta)\). The pdfs in (10), (11) and (12) are chosen because they combine nicely with the likelihood function for derivation of the conditional posterior pdfs, and because they are sufficiently flexible to represent vague prior information which can be dominated by the sample
data. In addition to the above prior pdfs, the restriction $\mu_1 < \mu_2 < \cdots < \mu_k$ is imposed a priori to ensure identifiability of the posterior distribution.

After completing the algebra necessary to combine the prior pdfs with the likelihood function in such a way that isolates the conditional posterior densities for use in a Gibbs’ sampler, we obtain the following conditional posterior pdfs.

The posterior probability that the $i$-th observation comes from the $z$-th component in the mixture, conditional on the unknown parameters, is the discrete pdf

$$P(Z_i = z | x, w, \pmb{\mu}, \pmb{\varphi}) = \frac{p_z}{p_{n1} + p_{n2} + \cdots + p_{nk}}$$

where

$$p_z = w_z \left( \frac{v_z / \mu_z}{\Gamma(v_z)} \right)^{v_z} x_i^{-v_z - 1} \exp \left\{ -\frac{v_z x_i}{\mu_z} \right\}$$

The posterior pdf for the mixture-component probabilities $w$, conditional on the other parameters and on the realized components for each observation $z = (z_1, z_2, \ldots, z_n)'$, is the Dirichlet pdf

$$f(w | x, \pmb{\mu}, \pmb{\varphi}) = D(n + \pmb{\varphi})$$

where $n = (n_1, n_2, \ldots, n_k)'$, with $n_z$ being the number of observations for which $Z_i = z$. Thus, $\sum_{z=1}^{k} n_z = n$.

The posterior pdfs for the means of the component densities $\mu_z$, conditional on the other parameters and on $z$, are the inverted gamma pdfs

$$f(\mu_z | x, z, w, \pmb{\varphi}) = GI(\alpha_z + n_z v_z, \beta_z + S_z v_z)$$

where $S_z = \sum_{i : Z_i = z} x_i$.

The form of the posterior pdfs for the scale parameters of the component densities $\nu_k$, conditional on the other parameters and on $z$, is not a common recognizable one. It is given by
where \( P_z = \prod_{i \in Z, z} x_i \).

A Gibbs sampling algorithm that iterates sequentially and iteratively through the conditional posterior pdfs can proceed as follows:

1. Set \( t = 0 \) and initial values \( \mathbf{w}^{(0)}, \mathbf{v}^{(0)} \).
2. Generate \( \mathbf{z}^{(t+1)} | \mathbf{x}, \mathbf{w}^{(t)}, \mathbf{v}^{(t)} \) from (6).
3. Generate \( \mathbf{w}^{(t+1)} | \mathbf{x}, \mathbf{z}^{(t+1)}, \mathbf{v}^{(t)} \) from (7).
4. Generate \( \mathbf{\mu}^{(t+1)} | \mathbf{x}, \mathbf{z}^{(t+1)}, \mathbf{v}^{(t)}, \mathbf{w}^{(t+1)} \) from (8), for \( z = 1, 2, \ldots, k \).
5. Generate \( \mathbf{v}^{(t+1)} | \mathbf{x}, \mathbf{\mu}^{(t+1)}, \mathbf{w}^{(t+1)} \) from (9), for \( z = 1, 2, \ldots, k \).
6. Order the elements for \( \mathbf{\mu}^{(t+1)} \) such that \( \mu_1 < \mu_2 < \ldots < \mu_k \) and sort \( \mathbf{w}^{(t+1)} \) and \( \mathbf{v}^{(t+1)} \) accordingly.
7. Set \( t = t + 1 \) and return to step 2.

To describe each of the generation steps in more detail, first consider (13). In this case we divide the interval \((0,1)\) into \( k \) sub-intervals with the length of the \( z \)-th sub-interval equal to \( P(Z_i = z | \mathbf{x}, \mathbf{w}, \mathbf{v}) \). A uniform random number is generated from the \((0,1)\) interval. The value assigned to \( Z_i \) is the sub-interval in which the uniform random number falls. To generate observations from the Dirichlet density in (14), we first generate \( k \) gamma random variables, say \( \gamma_z, z = 1, 2, \ldots, k \) from \( G(\phi_z + n_z, 1) \) densities, and then set \( w_z = \gamma_z / \sum_{j=1}^k \gamma_j \). To generate \( \mu_z \) from (15), we generate a random variable from a \( G(\alpha_z + n_z v_z, \beta_z + S_z v_z) \) density and then invert it.

Generating \( v_z \) from equation (16) is more complicated, requiring a Metropolis step. We draw a candidate \( \bar{v}_z^{(t+1)} \) from a gamma density with mean equal to the previous draw \( v_z^{(t)} \). That is, a candidate \( \bar{v}_z^{(t+1)} \) is generated from a \( G(r, r/v_z^{(t)}) \) distribution and is accepted as \( v_z^{(t+1)} \) with
probability
\[
\min \left\{ \frac{f(y_{z(t)}^{(t+1)} | x, z^{(t+1)}, w^{(t+1)}, v_{z(t)}^{(t+1)}, v_{z(t)}^{(t+1)})}{f(y_{z(t)}^{(t)} | x, z^{(t)}, w^{(t+1)}, v_{z(t)}^{(t+1)}, v_{z(t)}^{(t+1)})} p\left(y_{z(t)}^{(t+1)}, v_{z(t)}^{(t+1)}\right) \right\}
\]

where \( p(y_{z(t)}, v_{z(t)}^{(t+1)}) \) is the gamma density used to generate \( y_{z(t)}^{(t+1)} \). Non-acceptance of \( y_{z(t)}^{(t+1)} \) implies \( y_{z(t)}^{(t+1)} = y_{z(t)}^{(t)} \). The value of \( r \) is chosen by experimentation to give an acceptance rate of approximately 0.25 to 0.4.

4. Data and Estimation Results
The data used to illustrate the methodology are the income data obtained for Indonesian urban regions for the years 1999, 2002, 2005 and 2008. This data was created using the household expenditure data obtained from the National Socio-Economic Survey (Susenas). Appropriate price and equivalence-scale adjustments were made to ensure comparability over years. Summary statistics for the data are presented in Table 1. The units are thousands of rupiah per month.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample mean</td>
<td>332.63</td>
<td>432.36</td>
<td>477.04</td>
<td>454.08</td>
</tr>
<tr>
<td>Median</td>
<td>270.40</td>
<td>337.80</td>
<td>357.95</td>
<td>355.27</td>
</tr>
<tr>
<td>minimum</td>
<td>44.12</td>
<td>57.65</td>
<td>38.32</td>
<td>59.84</td>
</tr>
<tr>
<td>maximum</td>
<td>5973.92</td>
<td>24,902.67</td>
<td>30,216.51</td>
<td>13,181.99</td>
</tr>
<tr>
<td>standard deviation</td>
<td>249.32</td>
<td>477.29</td>
<td>511.19</td>
<td>401.23</td>
</tr>
<tr>
<td>sample size</td>
<td>25,175</td>
<td>29,280</td>
<td>24,687</td>
<td>26,648</td>
</tr>
<tr>
<td>proportion less than 60</td>
<td>0.00024</td>
<td>0.00007</td>
<td>0.00057</td>
<td>0.00004</td>
</tr>
<tr>
<td>proportion more than 2000</td>
<td>0.00242</td>
<td>0.0074</td>
<td>0.0134</td>
<td>0.0099</td>
</tr>
</tbody>
</table>

If one makes judgments on the ordering of the distributions on the basis of the means or medians, the population becomes better off as it moves from 1999 to 2005, but drops back in 2008, a likely consequence of the global financial crisis. Similarly, using the standard deviation to measure inequality suggests inequality increases from 1999 to 2005, but then decreases in 2008. Assuming a welfare function where mean income contributes positively and the standard
deviation contributes negatively, it is not clear whether the distribution in any one year is preferable, although 2008 appears to dominate 2002.

Comparing the mean and median incomes, and checking the mean against the maximum values, reveals an extremely long right tail, with a small number of households with very high incomes. For example, in 2002, the mean income was 432, the maximum income was 24,903, and the proportion of households with incomes greater than 2,000 was only 0.0074. Our dominance tests were performed for incomes in the interval (60, 2000). Table 1 also presents the proportions of sample observations that are omitted through the selection of this range.

For estimating the gamma mixtures, we tried 2, 3, 4 and 5 components and finally settled on 5 components. For 2 and 3 components we used an MCMC sample of 50,000 of which 5,000 were discarded as a burn in. For 4 and 5 components, we used an MCMC sample of 200,000 of which 50,000 were discarded as a burn in. To assess goodness-of-fit, we computed chi-square statistics and root mean square errors that compared the sample proportions of observations in each of 50 groups with those estimated from the gamma mixtures, with the parameters set equal to their posterior means. The values for these statistics for 2, 3, 4 and 5 components are presented in Table 2. There were dramatic improvements in these values up to 4 components, but only moderate improvements (and in some cases no improvement) when going from 4 to 5 components, and so we settled on 5 components. There are some chi-square values that are significantly different from zero at a 1% significance level, but they appear respectable, given the large sample sizes. Also, the plots of the predictive densities from 5 components against the histograms given in Figure 1 show that the essential shapes of the distributions have been captured.

<table>
<thead>
<tr>
<th>Table 2: Goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square values</td>
</tr>
<tr>
<td>k=2</td>
</tr>
<tr>
<td>1999</td>
</tr>
<tr>
<td>2002</td>
</tr>
<tr>
<td>2005</td>
</tr>
<tr>
<td>2008</td>
</tr>
</tbody>
</table>
Chi-square critical values are 67.505 at 5% and 76.154 at 1%

Figure 1(a) Histogram and Predictive Density for 1999

Figure 1(b) Histogram and Predictive Density for 2002
Posterior means and standard deviations for each of the parameters for all components and all years are given in Table 3. In general the weights for the last two components are very small. In fact the means for the last component are greater than 2000, the largest value in our graphs in Figure 1, and the upper limit we used for dominance assessment. We can likely attribute the component with the large mean to the long tails.
### Table 3: Posterior Summary Statistics for the Parameters

<table>
<thead>
<tr>
<th>Year</th>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>( \mu )</td>
<td>156.62</td>
<td>238.55</td>
<td>362.98</td>
<td>613.43</td>
<td>1586.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.0286)</td>
<td>(13.5186)</td>
<td>(23.4875)</td>
<td>(60.4521)</td>
<td>(268.9961)</td>
</tr>
<tr>
<td></td>
<td>( w )</td>
<td>0.1546</td>
<td>0.3186</td>
<td>0.3851</td>
<td>0.1316</td>
<td>0.0100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0608)</td>
<td>(0.0895)</td>
<td>(0.0687)</td>
<td>(0.0348)</td>
<td>(0.0031)</td>
</tr>
<tr>
<td></td>
<td>( v )</td>
<td>19.2094</td>
<td>13.3874</td>
<td>7.4097</td>
<td>4.8993</td>
<td>2.3627</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.3872)</td>
<td>(3.8888)</td>
<td>(1.3227)</td>
<td>(1.0648)</td>
<td>(0.7327)</td>
</tr>
<tr>
<td>2002</td>
<td>( \mu )</td>
<td>210.87</td>
<td>350.58</td>
<td>571.02</td>
<td>1089.62</td>
<td>5165.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.2107)</td>
<td>(16.6166)</td>
<td>(47.3702)</td>
<td>(80.4297)</td>
<td>(1022.167)</td>
</tr>
<tr>
<td></td>
<td>( w )</td>
<td>0.2664</td>
<td>0.4027</td>
<td>0.2579</td>
<td>0.0699</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0504)</td>
<td>(0.0762)</td>
<td>(0.0664)</td>
<td>(0.0119)</td>
<td>(0.0008)</td>
</tr>
<tr>
<td></td>
<td>( v )</td>
<td>11.8178</td>
<td>9.6136</td>
<td>7.7290</td>
<td>3.5955</td>
<td>1.3328</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.7002)</td>
<td>(1.7140)</td>
<td>(2.2162)</td>
<td>(0.5818)</td>
<td>(0.3933)</td>
</tr>
<tr>
<td>2005</td>
<td>( \mu )</td>
<td>219.65</td>
<td>385.76</td>
<td>648.90</td>
<td>1299.93</td>
<td>3426.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.4539)</td>
<td>(23.2643)</td>
<td>(72.4493)</td>
<td>(214.5909)</td>
<td>(654.5322)</td>
</tr>
<tr>
<td></td>
<td>( w )</td>
<td>0.2663</td>
<td>0.4228</td>
<td>0.2384</td>
<td>0.0649</td>
<td>0.0077</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0533)</td>
<td>(0.0823)</td>
<td>(0.0695)</td>
<td>(0.0225)</td>
<td>(0.0026)</td>
</tr>
<tr>
<td></td>
<td>( v )</td>
<td>9.9940</td>
<td>7.2201</td>
<td>5.6900</td>
<td>3.7374</td>
<td>1.5698</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.8013)</td>
<td>(1.3936)</td>
<td>(1.4589)</td>
<td>(1.7077)</td>
<td>(0.4409)</td>
</tr>
<tr>
<td>2008</td>
<td>( \mu )</td>
<td>188.31</td>
<td>322.74</td>
<td>543.50</td>
<td>1146.31</td>
<td>2853.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.2113)</td>
<td>(13.3876)</td>
<td>(30.5865)</td>
<td>(104.1423)</td>
<td>(585.2628)</td>
</tr>
<tr>
<td></td>
<td>( w )</td>
<td>0.1718</td>
<td>0.3788</td>
<td>0.3686</td>
<td>0.0746</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0290)</td>
<td>(0.0620)</td>
<td>(0.0543)</td>
<td>(0.0148)</td>
<td>(0.0026)</td>
</tr>
<tr>
<td></td>
<td>( v )</td>
<td>13.8416</td>
<td>8.9406</td>
<td>6.2767</td>
<td>3.8587</td>
<td>2.1298</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.1084)</td>
<td>(1.3487)</td>
<td>(0.7121)</td>
<td>(0.8558)</td>
<td>(0.6988)</td>
</tr>
</tbody>
</table>

### 5. Dominance Results

Given distributions for 4 years, there are 6 possible pairwise comparisons that can be made. For each comparison (say, \( A \) versus \( B \)), and each dominance criterion, there are 3 probabilities that can be computed: (1) the probability that \( A \) dominates \( B \), (2) the probability that \( B \) dominates \( A \), and (3) the probability that neither \( A \) nor \( B \) is dominant. These dominance probabilities are
reported in Table 4 for income values from 60 to 2000; the proportions of parameter draws satisfying the required inequalities were evaluated for income values within this range using increments of 1. The range (60, 2000) includes more than 98% of the population in each of the years. The sample proportions omitted from each tail are given in Table 1. For each pairwise comparison we report the probabilities for $A$ dominating $B$, and $B$ dominating $A$. The probability that neither $A$ nor $B$ dominates is given by $1 - \Pr(A \geq_D B) - \Pr(B \geq_D A)$. The diagonal entries in Table 4 provide evidence of how the income distribution has changed from one period to the next. With probabilities all greater than 0.96, there is strong evidence for all forms of dominance for 2002 over 1999, and zero probabilities for dominance in the other direction (bottom-right entry in the table). Moving to a comparison of 2005 and 2002, we find probabilities of 1 that neither year is dominant for FSD and SSD, and almost 1 for LD. There is a slight probability (0.00014) that 2005 Lorenz dominates 2002. We obtain a similar result when moving from 2005 to 2008. There are some very small probabilities that 2008 is worse than 2005 in terms of FSD and SSD, and a very small probability that 2008 Lorenz dominates 2005, but the probabilities for neither year dominating are essentially 1.

What is perhaps surprising is that the FSD and SSD probabilities for 2005 over 1999 are not larger. They are approximately 0.23, much lower than the corresponding probabilities of 0.97 for 2002 over 1999. Since the mean for 2005 is greater than that for 2002, we might expect the FSD and SSD dominance probabilities for 2005 over 1999 to be greater than those for 2002 over 1999. It turns out that the probabilities for 2005 over 1999 are very sensitive to the starting point for incomes. If the starting point for incomes was moved from 60 to 100, then it turns out that both dominance probabilities are equal to 1. A similar result, but one less dependent on the very small incomes, occurs for dominance of 2008 over 2002. In this case the overall FSD and SSD probabilities are close to zero, but beyond incomes of 300, they are equal to 1.
In addition to examining dominance over the complete range of the distribution we can consider restricted ranges that may be of interest. For example, if we are interested in how the poor have fared we can compare the lower parts of the distribution for different years. Table 5 contains dominance results for incomes between 60 and 200. In 2005, 200,000 rupiah/month was approximately equal to 70 cents U.S. per day. The proportions of the sample within this range
were 0.278 in 1999, 0.161 in 2002, 0.153 in 2005 and 0.163 in 2008. For this restricted range, all dominance results that involve the year 1999 were similar to those for the unrestricted range. This occurs when the minimum value for a “probability curve”, that plots the probability of the relevant inequality being satisfied against each income value, lies within the restricted interval. The probability of dominance will necessarily be less than or equal to the minimum value of the probability curve. For comparisons using 1999, the minima of the probability curves were always less than 200, and this was true for dominance in both directions. For comparisons that did not involve 1999, there were some differences because some minima occurred at values greater than 200. For 2008 dominating 2005, the probability of SSD increased to 0.15. Also, 2002 was better than both 2005 and 2008 in the sense that there were positive probabilities of FSD and SSD for 2002 over the later years.

6. Concluding Remarks

The development of statistical inference for assessing whether income distributions have changed over time in what might be considered a desirable way has attracted a great deal of attention within the sampling-theory framework. Hypothesis testing procedures have been developed for, among other things, Lorenz dominance, generalized Lorenz dominance and first-order stochastic dominance. The purpose of this paper was to illustrate how such dominance relationships can be assessed within a framework of Bayesian inference. Bayesian inference has the advantage of reporting results in terms of probabilities - a natural way to express our uncertainty. Because it enables us to give probabilities for dominance in either direction, as well as the probability that dominance does not occur, it overcomes the problem of giving favourable treatment to what is chosen as the null hypothesis in sampling theory inference.

References


