aix-marseille school of economics
école d'économie d'aix-marseille

Working Papers / Documents de travail

The Bayesian approach to poverty measurement

Michel Lubrano
Zhou Xun

WP 2021 - Nr 33

The Bayesian approach to poverty measurement*

Michel Lubrano[†] Zhou Xun[‡]

May 21, 2021

Abstract

This survey paper reviews the recent Bayesian literature on poverty measurement. After introducing Bayesian statistics, we show how Bayesian model criticism could help to revise the international poverty line. Using mixtures of lognormals to model income, we derive the posterior distribution for the FGT, Watts and Sen poverty indices, then for TIP curves (with an illustration on child poverty in Germany) and finally for Growth Incidence Curves. The relation of restricted stochastic dominance with TIP and GIC dominance is detailed with an example on UK data. Using panel data, we show how to decompose poverty into total, chronic and transient poverty, comparing child and adult poverty in East Germany when redistribution is introduced. When a panel is not available, a Gibbs sampler is used to build a pseudo panel. We illustrate poverty dynamics by examining the consequences of the Wall on poverty entry and poverty persistence in occupied West Bank.

Keywords: Bayesian inference, mixture model, poverty indices, stochastic dominance, poverty dynamics.
JEL codes: C11, C46, I32, I38.

*This work was supported by the French National Research Agency Grant ANR-17-EURE-0020, and by the Excellence Initiative of Aix-Marseille University - A*MIDEX. The support of the National Natural Science Foundation of China, Grant number 71764008 is also acknowledged.

[†]School of Economics, Jiangxi University of Finance and Economics, China and Aix Marseille Univ, CNRS, AMSE, Marseille, France

[‡]School of Economics, Nanjing University of Finance and Economics, China

1 Introduction

For long, standard errors were not reported for poverty or inequality indices, and this on two grounds. Data sets based on surveys included more than five thousands observations, so it was thought that the standard errors would have been very small. A second objection was the difficulty of computation (see for instance Davidson 2009 for the Gini index or Biewen and Jenkins 2006 for generalized entropy indices and complex sampling). These arguments are no longer tenable. We might well be interested in sub-groups, operating thus on reduced sample sizes. The Bayesian approach brings in feasible answers for small sample sizes and its simulation techniques make simple the computation of standard errors. More precisely, a Bayesian approach to poverty measurement relies most of the time on a parametric modelling of the income distribution. Poverty indices, the TIP curve of Jenkins and Lambert (1997), the growth incidence curve of Ravallion and Chen (2003) are transformations of the parameters of this parametric income distribution. The purpose of Bayesian inference will be to provide draws of the posterior density of these quantities, using simulation methods. The same approach will be used to explore restricted stochastic dominance and poverty dynamics. Before covering these applications, let us introduce what is Bayesian inference and what are the simulation methods involved.

1.1 Bayesian statistics

Let us consider a continuous random variable X. Realizations of this random variable will be our sample space and we shall assume that it is equipped with a particular structure of σ-field so that we can define probability measures over it. These measures are indexed by a parameter θ belonging to a parameter space Θ. If we assume that Θ is dominated by a $\sigma-$ finite measure (this is a restriction), probabilities over the sample space X can be described by a density function $p(x|\theta)$, leaving aside a non-parametric approach. Given N realizations $(x_1, x_2, \ldots, x_i, \ldots, x_N)$, we can write the likelihood function of this observed sample:

$$\ell(\theta; x) = \prod_{i=1}^{N} p(x_i|\theta).$$

Classical inference is looking for the value of θ that is the most likely to have produced the observed sample, assuming that given the data density, there is somewhere a true value of the parameter on which we want to get information given the observed sample. However, as there might exist other realizations of X, one of the main concerns of classical statisticians is to investigate: *What would happen if our sample size were tending to infinity?* *Bayesian inference* follows a different route. We equip the parameter space Θ with a probability structure so that a prior probability $\varphi(\theta)$ can be defined. This means that there is no unique true value of θ, but uncertainty around possible values that θ could take. The object of inference is to reduce this uncertainty by learning from the observation of a realization x of X, using Bayes' theorem:

$$\varphi(\theta|x) = \frac{\ell(\theta; x) \times \varphi(\theta)}{p(x)}. \tag{1}$$

$\varphi(\theta)$ is the prior density of θ which describes our prior knowledge around the plausible values of θ. $\ell(\theta; x)$ is the likelihood function of the sample and the common element with the classical approach. $\varphi(\theta|x)$ is the posterior density of θ, which means how our prior knowledge of θ was revised by the observation of one realization x of the random variable X. By realization, we mean the observation of a sample of a given size. Following Lindley (1971), the purpose of Bayesian inference is quite different from the classical approach: once a sample is realized, it is no longer considered as a random variable. Bayes theorem is a learning mechanism, describing how a prior knowledge is revised by the experience of observing the realization of a particular sample. Finally in (1), $p(x)$ is the predictive density:

$$p(x) = \int \ell(\theta; x) \times \varphi(\theta)\, d\theta. \tag{2}$$

It gives the probability of observing a particular realization of our sample x, given all the possible prior values of the parameter θ. Its evaluation is rarely necessary. If $\ell(\theta; x)$ and $\varphi(\theta)$ belong to well-known families, the integrating constant of the posterior density $\varphi(\theta|x)$ can be recovered analytically. So Bayes theorem in this case can be simplified to:

$$\varphi(\theta|x) \propto \ell(\theta; x) \times \varphi(\theta),$$

which means that the posterior density is proportional to the product of the likelihood function times the prior.

1.2 Simulation methods

For models more complex than the linear regression model, it is not possible to find an analytical expression for $\varphi(\theta|x)$. However in many cases, it is possible to decompose θ in several blocks, say $\theta' = [\theta'_1, \theta'_2]$ and to recover analytical expressions for $\varphi(\theta_1|\theta_2, x)$ and $\varphi(\theta_2|\theta_1, x)$ from which it is possible to draw random numbers. In this case, the alternative sampling from:

$$\theta_1^{(j)} \sim \varphi(\theta_1|\theta_2^{(j-1)}, x), \tag{3}$$
$$\theta_2^{(j)} \sim \varphi(\theta_2|\theta_1^{(j)}, x), \tag{4}$$

produces a Markov chain that converges under mild conditions to draws from the posterior distribution $\varphi(\theta|x)$. The method is called a Gibbs sampler and belongs to the class of Monte Carlo Markov Chains or MCMC. When this type of decomposition is not possible, one has to rely on a Metropolis algorithm and an importance function from which it is easy to draw random numbers and which is not too far from $\varphi(\theta|x)$.

2 Revising the IPL using Bayesian inference

Using Bayesian inference, we revise the international poverty line (IPL) of the World Bank which serves to count the number of poor in the world and locate poverty in order to design anti-poverty policies.

2.1 The econometric model of the World Bank

The international poverty line of the World Bank relies from a constrained regression model and a data base covering 74 developing countries. Ravallion et al. (2009) note that below a certain level of consumption, national poverty lines z_i seem to be constant while they evolve as a function of consumption after that level:

$$z_i = s_i(\alpha_1 + \gamma_1 C_i) + (1 - s_i)(\alpha_2 + \gamma_2 C_i) + \epsilon_i, \tag{5}$$

where s_i is equal to an indicator function $\mathbb{I}(C_i < \theta)$, which is one for countries below a mean consumption of θ and zero otherwise. For $C_i < \theta$, the constraint $\gamma_1 = 0$ is imposed, corresponding to the concept of an absolute poverty line. Moreover, Ravallion et al. (2009) do not estimate θ, but fix it to \$60 per month. With all these restrictions, the IPL corresponds to the estimated value of α_1 and is found to be \$1.25 per day when using 2005 PPP.

2.2 Bayesian model criticism: Poverty and social inclusion

Model criticism would imply estimating a complete switching regression model where at least θ is unknown. However, classical inference is not well suited in this case as according to Hansen (2000), the asymptotic distribution of the threshold parameter is non-standard and can be misleading in small samples. A more intuitive alternative is to apply the Bayesian approach which is by nature more robust for small sample sizes and provides a direct inference process for the posterior distribution of θ.

Xun and Lubrano (2018) enlarge the initial model (5) by introducing the notion of social inclusion developed in Atkinson and Bourguignon (2001). At any level of income, poverty corresponds to the deprivation of enough resources to participate in social life. Social inclusion in Atkinson and Bourguignon (2001) means that poverty is not only a matter of minimum calory consumption (absolute poverty line), but also depends on social life participation. As a measure of social inclusion, Xun and Lubrano (2018) consider the unemployment rate ur_i, leading to the richer econometric model:

$$z_i = s_i(\alpha_1 + \gamma_1 \log C_i + \beta_1 ur_i) + (1 - s_i)(\alpha_2 + \gamma_2 \log C_i) + \epsilon_i, \tag{6}$$

$$s_i = \mathbb{I}(C_i < \theta), \tag{7}$$

$$\mathrm{Var}(\epsilon_i) = s_i\sigma_1^2 + (1 - s_i)\sigma_2^2. \tag{8}$$

This is a switching regression model with heteroskedasticity where θ is an unknown parameter. Bayesian inference provides posterior draws for the parameters, leading to a much larger definition for the group of developing countries since $\mathrm{E}(\theta|z) = 169.2$ (14.03), as illustrated in the left panel of Figure 1. We have now 39 countries in that group instead of 15 in Ravallion et al. (2009). Posterior draws are then converted into posterior draws for the poverty line using:

$$z^{(j)} = \frac{1}{n_j} \sum_i (\gamma_1^{(j)} \log(C_i) + \beta_1^{(j)} ur_i) \mathbb{I}(C_i < \theta^{(j)}) \tag{9}$$

where n_j is the number of observations in the first regime given the j^{th} draw (α_1 was discarded as it was not significantly different from 0). The posterior expectation of the IPL is found to be \$1.48 (0.036), a rather greater value than the \$1.25 IPL of the World Bank which by the way does not belong to the credible HPD region of 90% [1.30, 1.65], leading thus to a substantial and significant revision as displayed in the right panel of Figure 1.

3 Poverty indices and poverty curves

Poverty indices are a way to summarize the left tail of an income distribution $f(x)$, obeying different types of axioms (see e.g. the survey of Zheng 1997).

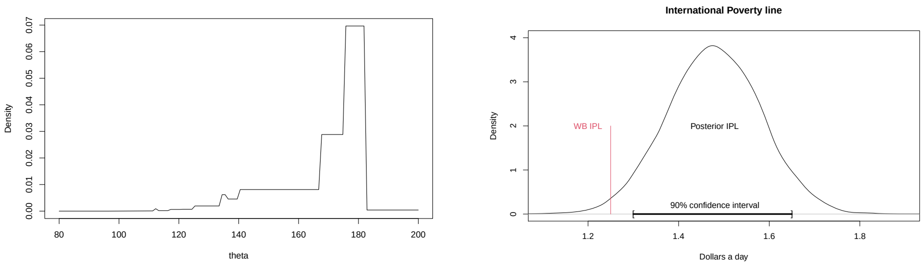


Figure 1: Posterior distribution of θ and of the IPL

Poverty indices are thus particular transformations of the income distribution. In a Bayesian Bayesian framework, the usual route is to consider a parametric model $f(x|\theta)$ for the income distribution. Once we have obtained draws from the posterior distribution of θ, we can transform these draws into draws of various poverty indices. Because there is no universal rule for selecting a particular poverty index, Jenkins and Lambert (1997) introduced TIP curves which document at the same time the three dimensions of poverty for each quantile of $f(x|\theta)$. Later Ravallion and Chen (2003) considered that growth is favourable to the poor if the lower quantiles of $f(x|\theta)$ increase more than its higher quantiles.

3.1 Modelling the income distribution

A mixture of distributions accounts for the fact that the population is made of different groups with specific characteristics while the belonging to a particular group is not observed. The general formulation of a finite mixture with K members is:

$$f(x|\theta) = \sum_{k=1}^{K} \eta_k f_k(x|\theta_k),$$

where η_k are the weights summing to 1 and θ_k the parameters of each member. Mixtures have very nice properties due to their linearity. In particular, the mean and the cumulative distribution (CDF) have direct expressions with:

$$\mathrm{E}(x|\theta) = \sum_{k=1}^{K} \eta_k \int_0^{\infty} x \, f_k(x|\theta_k) \, dx,$$

and

$$F(x|\theta) = \sum_{k=1}^{K} \eta_k F_k(x|\theta_k).$$

So the mean is weighted average of the mean of each component and the CDF is the weighted average of each component CDF. Various choices have been made in the Bayesian literature. Gunawan et al. (2020) used a mixture of three gamma densities for Australia to evaluate the posterior distribution of a head count index. They study the impact of using or not sampling weights and show that it leads to different evaluations of poverty. Ndoye and Lubrano (2014) use a mixture of two Pareto distributions to analyse top wage inequality in the US. Lubrano and Ndoye (2016) opting for a mixture of lognormals derive the posterior density of a Gini inequality index and detail the decomposition of the Generalized Entropy index. In this paper, we review the choice of modelling $f(x|\theta)$ as a mixture of lognormal densities.

The lognormal density is noted:

$$f_\Lambda(x|\theta) = \frac{1}{x\sigma\sqrt{2\pi}} \exp -\frac{(\log x - \mu)^2}{2\sigma^2}, \tag{10}$$

with CDF:

$$F_\Lambda(x|\theta) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \tag{11}$$

where Φ is the Gaussian CDF. The mean and variance are:

$$\mathrm{E}(x|\theta) = e^{\mu+\sigma^2/2}, \quad \mathrm{Var}(x|\theta) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}. \tag{12}$$

The first partial moments are (see e.g. Jawitz 2004):

$$\int_0^z x f_\Lambda(x|\theta) = e^{\mu+\sigma^2/2}\Phi\left(\frac{\log(z) - \mu - \sigma^2}{\sigma}\right) \tag{13}$$

$$\int_0^z x^2 f_\Lambda(x|\theta) = e^{2\mu+2\sigma^2}\Phi\left(\frac{\log(z) - \mu - 2\sigma^2}{\sigma}\right) \tag{14}$$

Mixtures of distributions are usually estimated using a Gibbs sampler, considering a mixture as an incomplete data problem. An auxiliary integer variable ζ allocates each observation x_i to a member of the mixture, identified by its label so that conditionally on a given sample allocation $[\zeta_i = k]$, each component of the mixture can be analysed separately using a natural conjugate prior. An algorithm is detailed in Lubrano and Ndoye (2016) while Fourrier-Nicolaï and Lubrano (2020) consider the case of sampling weights and zero incomes (see also Gunawan et al. 2020). The posterior distribution of various poverty indices can be obtained as transformations of the m draws collected in the Gibbs output, indexed by j. From these m draws, we can compute a mean, a standard error, a posterior confidence interval and plot the posterior density.

3.2 Posterior draws for poverty indices

The general class of poverty indices of Foster et al. (1984) writes:

$$FGT(z, \alpha) = \int_0^z (1 - x/z)^\alpha f(x)\, dx, \tag{15}$$

The poverty head-count ratio or poverty rate H corresponds to $\alpha = 0$. It leads to a simple solution:

$$H(z|\theta^{(j)}) = \sum_{k=1}^{K} \eta_k^{(j)} \Phi\left(\frac{\log z - \mu_k^{(j)}}{\sigma_k^{(j)}}\right). \tag{16}$$

For $\alpha = 1$, the poverty gap index can be decomposed into:

$$\int_0^z (1 - x/z) f(x) dx = F(z) - \frac{1}{z}\int_0^z x f(x)\, dx.$$

Using (11) and (13), we have:

$$\begin{aligned} FGT(z|\theta^{(j)}, 1) &= \sum_{k=1}^{K} \eta_k^{(j)} \left[\Phi\left(\frac{\log z - \mu_k^{(j)}}{\sigma_k^{(j)}}\right)\right. \\ &\quad \left. - \frac{1}{z} e^{\mu_k^{(j)} + \sigma_k^{2(j)}/2} \Phi\left(\frac{\log z - \mu_k^{(j)} - \sigma_k^{2(j)}}{\sigma_k^{(j)}}\right)\right]. \end{aligned} \tag{17}$$

For $\alpha = 2$, we have to evaluate

$$\int_0^z f(x)\, dx - \frac{2}{z}\int_0^z x f(x)\, dx + \frac{1}{z^2}\int_0^z x^2 f(x)\, dx.$$

Using (11), (13), (14), we get:

$$\Phi\left(\frac{\log z - \mu}{\sigma}\right) - \frac{2}{z} e^{\mu + \sigma^2/2} \Phi\left(\frac{\log(z) - \mu - \sigma^2}{\sigma}\right) + \frac{1}{z^2} e^{2\mu + 2\sigma^2} \Phi\left(\frac{\log(z) - \mu - 2\sigma^2}{\sigma}\right),$$

so that for a mixture of lognormals we have:

$$\begin{aligned} FGT(z|\theta^{(j)}, 2) &= \sum_{k=1}^{K} \eta_k^{(j)} \left[\Phi\left(\frac{\log x - \mu_k^{(j)}}{\sigma_k^{(j)}}\right) - \frac{2}{z} e^{\mu + \sigma_k^{2(j)}/2} \Phi\left(\frac{\log z - \mu_k^{(j)} - \sigma_k^{2(j)}}{\sigma_k^{(j)}}\right)\right. \\ &\quad \left. + \frac{1}{z^2} e^{2\mu_k^{(j)} + 2\sigma_k^{2(j)}} \Phi\left(\frac{\log(z) - \mu_k^{(j)} - 2\sigma_k^{2(j)}}{\sigma_k^{(j)}}\right)\right]. \end{aligned} \tag{18}$$

The Watts (1968) poverty index writes:

$$W(z) = -\int_0^z \log(x/z) f(x)\, dx. \tag{19}$$

Muller (2001) gave its expression when $f(x)$ is a lognormal:

$$W(z) = (\log z - \mu)\Phi\left(\frac{\log z - \mu}{\sigma}\right) + \sigma\phi\left(\frac{\log z - \mu}{\sigma}\right),$$

where ϕ is the Gaussian probability density. The generalization to mixtures provides:

$$W(z|\theta^{(j)}) = \sum_{k=1}^K \eta_k^{(j)} \left[(\log z - \mu_k^{(j)})\Phi\left(\frac{\log z - \mu_k^{(j)}}{\sigma_k^{(j)}}\right) + \sigma\phi\left(\frac{\log z - \mu_k^{(j)}}{\sigma_k^{(j)}}\right)\right].$$

The revision of Sen index by Shorrocks (1995) leads to:

$$SST(z) \quad = \quad \frac{2}{z}\int_0^z (z - x)(1 - F(x)) f(x)\, dx, \tag{20}$$

as expressed in Davidson (2009). We can decompose it into:

$$SST(z)/2 = FGT(z, 1) - \int_0^z (z - x) F(x) f(x)\, dx.$$

The last integral is related to the Gini index and has no analytical solution. In a similar situation, Lubrano and Ndoye (2016) proposed to evaluate numerically the integral for each draw of the parameters, using a Simpson rule.

3.3 TIP curves

The TIP curve of Jenkins and Lambert (1997) documents the three dimensions of poverty for each quantile of the income distribution up to the quantile corresponding to the poverty line z:

$$TIP(p, z) = \int_0^{F^{-1}(p)} (1 - x/z)\mathbb{I}(x \leq z) f(x)\, dx.$$

Letting $q = F^{-1}(p)$, we can decompose this equation into:

$$TIP(p, z) = \int_0^q f(x)\, dx - \frac{1}{z}\int_0^q y f(x)\, dx = p - \frac{1}{z}GL(p), \qquad \text{for } p \leq F(z),$$

where $GL(p)$ is the generalized Lorenz curve. The whole expression has an analytical form for the lognormal distribution. But this is of little use as it is not possible to find the closed expression of $GL(p)$ when $f(x)$ is a mixture. So it is better to consider directly:

$$TIP(p,z) = \sum_{k=1}^{K} \eta_k \int_0^q f_\Lambda(x|\mu_k, \sigma_k^2) - \frac{1}{z} \sum_{k=1}^{K} \eta_k \int_0^q x f_\Lambda(x|\mu_k, \sigma_k^2)\, dx,$$

where the quantile q has to be calculated separately. This presentation relies on the two-equation definition of the Lorenz curve, in use before Gastwirth (1971). Both integrals have an analytical solution leading to:

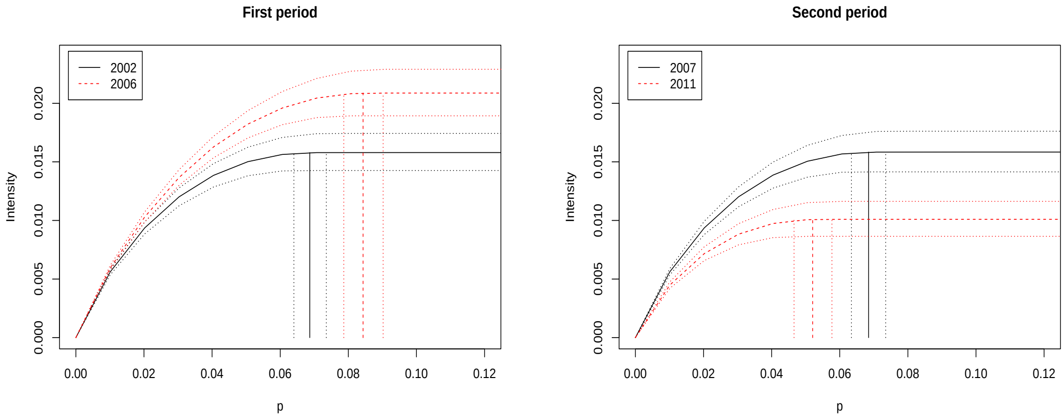
$$\begin{aligned} TIP(p,z|\theta^{(j)}) \quad = \quad & \sum_{k=1}^{K} \eta_k^{(j)} \left[\Phi\left(\frac{\ln q^{(j)} - \mu_k^{(j)}}{\sigma_k^{(j)}} \right) \right. \\ & \left. - \quad \frac{1}{z} e^{\mu_k^{(j)} + \sigma_k^{2(j)}/2} \Phi\left(\frac{\ln q^{(j)} - \mu_k^{(j)} - \sigma_k^{2(j)}}{\sigma_k^{(j)}} \right) \right]. \end{aligned} \tag{21}$$

The difficulty is that the left-hand side is a function of p while the right-hand side is a function of q. For each draw of θ, we have to solve numerically the equation:

$$F(q^{(j)}|\theta^{(j)}) = p, \tag{22}$$

for each point of a predefined grid on p. This is a feasible problem because it is of dimension one on a finite interval defined by the range of x. Brent (1971) algorithm is very efficient in this case. Note that Lander et al. (2020) advocate a different approach: For each draw of the MCMC output, they generate a vector of incomes from the posterior predictive distribution $p(x|\theta^{(j)})$ and find for a given grid of p the empirical quantile function, using interpolation.

The feasibility of the method is illustrated in Figure 2, extracted from Fourrier-Nicolaï and Lubrano (2020). It depicts the evolution of child poverty in Germany between 2002 and in 2011. The period has experienced a dramatic change in family social allowances. In each panel, vertical lines represent poverty headcount, horizontal lines poverty intensity and the curvature of the TIP curves poverty inequality. The fact that credible intervals do not overlap indicates that child poverty has significantly changed over the period. It increased a lot between 2002 and 2006 to finally decrease between 2007 and 2011. The change in family social policy has managed to cut the regular increase in child poverty that was documented in Corak et al. (2008).



90% credible intervals are represented by dotted lines. In black solid line is represented the TIP curve at the beginning of each sub-period. The red dashed line corresponds to the TIP curve of the end of each subperiod.

Figure 2: The three I's of child poverty in Germany

3.4 Pro-poor growth

The Growth Incidence Curve (GIC) of Ravallion and Chen (2003) can be approximated by the difference between the logs of two quantile functions:

$$g_t(p) = \log Q_t(p|\theta_t) - \log Q_{t-1}(p|\theta_{t-1}). \tag{23}$$

Because the quantile function corresponds to the first derivative of the generalized Lorenz curve, Fourrier-Nicolaï and Lubrano (2021) proposed two alternative ways for finding a parametric formulation for the GIC curve. The first method relies on finding the quantile function associated to a mixture of lognormal distributions. This requires solving (22) as seen above. The second method uses a direct modelling of the Lorenz curve. Several parametric forms were proposed in the literature, using one parameter (Chotikapanich 1993), two parameters (Kakwani and Podder 1973) or three parameters with Villasenor and Arnold (1989) or Kakwani (1980). The latter is built around the Beta density with:

$$L(p|\alpha) = p - \alpha_0 p^{\alpha_1}(1-p)^{\alpha_2}, \tag{24}$$

leading to the quantile function:

$$Q(p|\alpha) = \bar{y} \times (1 - \alpha_0\alpha_1 p^{\alpha_1 - 1}(1-p)^{\alpha_2} + \alpha_0\alpha_2 p^{\alpha_1}(1-p)^{\alpha_2 - 1}).$$

Bayesian inference on the parameters of (24) is obtained by considering the linear regression:

$$\log(p_i - \hat{L}_i) = \log(\alpha_0) + \alpha_1 \log(p_i) + \alpha_2 \log(1-p_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \tag{25}$$

where $\hat{L}_i = L(p_i = i/n) = \sum_{j=1}^{i} y[j]/\bar{y}$, $y[j]$ being the order statistics. Obtaining random draws from this quantile function requires some care as:

$$\begin{aligned} Q(p|\alpha^{(j)}, y) &= \bar{y}\exp(u^{(j)}) \times (1 - \exp(\alpha_0^{(j)})\alpha_1^{(j)} p^{\alpha_1^{(j)}-1}(1-p)^{\alpha_2^{(j)}} \\ &\quad + \exp(\alpha_0^{(j)})\alpha_2^{(j)} p^{\alpha_1^{(j)}}(1-p)^{\alpha_2^{(j)}-1}), \quad u^{(j)} \sim N(0, \sigma^{2(j)}). \end{aligned}$$

4 Restricted stochastic dominance

With restricted stochastic dominance, we can compare two poverty situations, whatever the social welfare function. By restricted stochastic dominance, we mean that two income distributions are compared up to a common poverty line (Davidson and Duclos 2000, 2013). Lander et al. (2020) use mixtures of gamma densities to model the income distribution in Indonesia. They develop Bayesian tests of stochastic dominance and restricted stochastic dominance. They compute posterior probabilities for stochastic dominance for the poorest 10% of the population, to assess whether their situation has improved over time. Comparing TIP curves is another way for testing restricted stochastic dominance at the second order while comparing two GICs relates to first order stochastic dominance.

Hypothesis testing is the domain where there is the greatest difference between classical and Bayesian approaches. Essentially in a Bayesian framework there is no privileged hypothesis. Two hypotheses H_0 and H_1 are compared by means of the ratio between their posterior probability, the famous Bayes factor $\Pr(H_0|y)/\Pr(H_1|y)$. We can also simply compute the posterior probability of H_0.

4.1 TIP dominance

TIP dominance compares two TIP curves defined for populations A and B.

Definition 1 *Distribution A TIP dominates distribution B for a given poverty line z if* $TIP_A(p, z) \leq TIP_B(p, z), \forall p \in [0, F(z)]$.

As underlined in Davidson and Duclos (2000), TIP dominance is related to restricted second order stochastic dominance. Testing for TIP dominance in a Bayesian frameworks leads first to compute for each draw of θ a vector $\delta(p|\theta)$ of dimension S corresponding to the grid over p:

$$\delta(p|\theta) = TIP_A(p, z|\theta_A) - TIP_B(p, z|\theta_B).$$

The condition $\delta(x, p|\theta) \leq 0$ defines a logical vector of zeros and ones. It is then equivalent to check any of the three conditions:

$$\prod_{i=1}^{S} \mathbb{I}[\delta(p_i|\theta) < 0] = 1, \quad \max_i \mathbb{I}[\delta(p_i|\theta) < 0] = 1, \quad \min_i \mathbb{I}[-\delta(p_i|\theta) > 0] = 1.$$

So for instance:

$$\begin{aligned} \Pr\left(\max_p d(p|y) < 0\right) &= \int_\theta \mathbb{I}\left[\max_p \delta(p|\theta) < 0\right] \varphi(\theta|y) d\theta \\ &\simeq \frac{1}{m}\sum_{j=1}^{m} \mathbb{I}\left[\max_p \delta(p|\theta^{(j)}) < 0\right], \end{aligned} \tag{26}$$

The range of p has to be slightly restricted because all TIP curves are zero at $p = 0$. So the practical range for the test should be something like $p \in [0.01, F(z)]$, values adopted in e.g. Davidson and Duclos (2013).

Because TIP dominance corresponds to restricted second order stochastic dominance, TIP dominance does not imply less poverty incidence. Using (16), we have to check the additional condition $H(z|\theta_A^{(j)}) < H(z|\theta_B^{(j)})$ and evaluate the proportion of draws when it is verified.

Finally, when can we say that the situation in A is not statistically different from the situation in B? Equality is rejected if, for at least one value of p_s, $\delta(p_s|\theta)$ is statistically different from zero. This means that we have to compute a credible interval for $\delta(p_s|\theta)$ and see if zero is included in this interval. If we find a single p_s for which zero does not belong to a say 90% credible interval for $\delta(p_s|\vartheta)$, then we can reject at the 90% level that the two TIP curves are equal.

4.2 GIC dominance

Because a GIC represents the difference between two quantiles functions, it corresponds to the p-approach to dominance of Davidson and Duclos (2000). We have first-order stochastic if $g_t(p) > 0$ for all p. Growth has been welfare-improving in terms of first-order stochastic dominance if $g_t(p) > 0$ for all p. We have restricted stochastic dominance if the range of p is limited to $p \in [0, F(z)]$. For each point p of a grid, Fourrier-Nicolaï and Lubrano (2021) evaluate:

$$\Pr(g_t(p) > 0) \simeq \frac{1}{m}\sum_{j=1}^{m} \mathbb{I}[g_t(p|\theta^{(j)}) > 0], \tag{27}$$

which allows us to see for which part of the income distribution the situation has been improved. The probability of dominance defined as:

$$\Pr(g_t(p) > 0) \simeq \frac{1}{m} \sum_{j=1}^{m} \mathbb{I}[\min_p (g_t(p|\theta^{(j)})) > 0]. \tag{28}$$

A further requirement is that growth has been favourable to the poor, leading to the vector corresponding to $p \in [0, F(z)]$:

$$\Pr(g_t(p) > \gamma) \simeq \frac{1}{m} \sum_{j=1}^{m} \mathbb{I}[g_t^{(j)}(p) > \gamma^{(j)}], \tag{29}$$

where

$$\gamma^{(j)} = \log \sum_k \eta_{k,2}^{(j)} e^{\mu_{k,2}^{(j)} + \sigma_{k,2}^{2(j)}} - \log \sum_k \eta_{k,1}^{(j)} e^{\mu_{k,1}^{(j)} + \sigma_{k,1}^{2(j)}} \tag{30}$$

is the j^{th} draw of the average growth rate between t and $t-1$ when the two income distributions are modelled as a mixture of lognormals.

Fourrier-Nicolaï and Lubrano (2021) analysed the impact of economic growth in the UK over the period 1979-1996 under the government of Margaret Thatcher. Using the Family Expenditure Survey, they found that

Table 1: Probability of pro-poor growth

p	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
					1979-1988				
$\Pr(g_t(p) > \gamma)$	0.00	0.00	0.00	0.00	0.00	0.01	0.29	0.93	1.00
					1992-1996				
$\Pr(g_t(p) > \gamma)$	1.00	1.00	1.00	0.99	0.93	0.77	0.58	0.39	0.19

growth has been profitable to the very top quantiles between 1979-1988. The next period 1992-1996 experienced strong fiscal and redistributive corrections leading to a situation which was more favourable to the lower quantiles. An approach using Kakwani modelling of the Lorenz curve did not give fundamentally different results.

5 Poverty dynamics

Hasegawa and Ueda (2007) propose to model individual incomes as a stationary process and derive the distribution of Ravallion (1988) decomposition of poverty into total, chronic and transitory poverty, using panel data.

However, panel data sets are seldom available in developing countries where the analysis of poverty should be of prime importance. Sadeq and Lubrano (2018) develop a pseudo panel approach to analyse the impact of the Wall on poverty entry and poverty persistence in the West Bank.

5.1 Poverty decomposition

TIP curves are a convenient graphical device to represent the three dimensions of poverty, thanks to the decomposability of FGT indices. When a panel data is available, a further decomposition is possible with total, transient and chronic poverty, following Ravallion (1988). Let y_{it} be income for individual i at time t. Hasegawa and Ueda (2007) assume that:

$$y_{it} = \mu_i + u_{it}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{31}$$

where μ_i represents the steady-state or long term income while u_{it} denotes its transient component. For FGT poverty indices expressed as the discrete counterpart of (15):

$$\pi(y_{it}, z) = \sum_{i=1}^{n} (1 - y_{it}/z)^{\alpha} \mathbb{1}(y_{it} < z),$$

total, chronic and transient poverty are measured by:

$$\begin{aligned}
\text{Total poverty} \quad \pi_F(z) &= \frac{1}{T} \sum_t \pi(y_{it}, z) && (32)\\
\text{Chronic poverty} \quad \pi_C(z) &= \frac{1}{n} \sum_i (1 - \mu_i/z)^{\alpha} \mathbb{1}(y_{it} < z) && (33)\\
\text{Transient poverty} \quad \pi_T(z) &= \pi_F(z) - \pi_C(z). && (34)
\end{aligned}$$

To go from a descriptive point of view to an inferential point of view, Hasegawa and Ueda (2007) model income by a mixture of k lognormal distributions for each individual i, assuming μ_i constant over time, but adding an error-in-variable mechanism. They derive the posterior predictive distribution of y_{it}, $p(\tilde{y}|y)$ and use simulations of $\tilde{y}$ to estimate poverty indices with $\hat{\mu}_i = \sum_t \tilde{y}_{it}/T$.

An alternative possibility would be to consider (31) as a panel data model with random individual effects. Let us define the vector of observations for an individual $y_i = [y_{i1}, \ldots, y_{iT}]$, the basic panel data model with random effects of Chib (1996) writes:

$$\begin{aligned}
y_i &= \iota\mu_i + X_i\beta + u_i, \qquad u_i|\sigma^2 \sim N(0, \sigma^2 I_T), && (35)\\
\mu_i &\sim N(0, \omega^2), && (36)
\end{aligned}$$

where ι is a vector of T ones. With a common random effect μ_i, the T incomes of individual i become correlated around the individual effect with:

$$\text{Var}(y_i|\beta, \sigma^2, \omega^2) = \sigma^2 I_T + \iota\iota'\omega^2 = V,$$

so that:

$$y_i \sim N(X_i\beta, V).$$

Bayesian inference on β, σ^2 and ω^2 is obtained with a Gibbs sampler corresponding to algorithm 2 of Chib and Carlin (1999) with an informative prior on σ^2 and ω^2 to ease convergence. With a MCMC output for $\beta^{(j)}$, $\sigma^{2(j)}$ and $\omega^{(j)}$, we can simulate m random draws for y_i using:

$$y_i^{(j)} \sim N(X_i\beta^{(j)}, \sigma^{2(j)} I_T + \iota\iota'\omega^{2(j)}).$$

We then transform each nT vector $y^{(j)} = [y_i^{(j)}]$ together with the by-product $\mu_i^{(j)}$ into:

$$\pi_F^{(j)}(z) = \frac{1}{nT}\sum_{i,T}(1 - y_{it}^{(j)}/z)^\alpha \mathbb{I}(y_{it}^{(j)} < z), \tag{37}$$

$$\pi_C^{(j)}(z) = \frac{1}{n}\sum_{i}(1 - \mu_i^{(j)}/z)^\alpha \mathbb{I}(\mu_i^{(j)} < z). \tag{38}$$

We have thus m posterior draws of the three poverty indices and compute standard error for each of them.

5.2 Child and adult poverty in East Germany

Using the data set of Fourrier-Nicolaï and Lubrano (2020), we analyse how social transfers were alleviating child poverty compared to adult poverty in East Germany over the period (2002-2006), just before the most important social and redistributive reforms introduced by the Hartz plan in 2006. We consider both disposable and market incomes (after taxes and transfers including family allowances or before taxes and allowances, divided by the new OECD equivalence scale) to build a five year balanced panel. We had 500 children and 1 466 adults without children. The poverty line is defined as 50% of the corresponding median income.

We adjusted a panel data model on the log of the income-to-need ratio, explained by an intercept, the household size and the number of children in the household (without the number of children for the adult sample). We used an informative inverted gamma2 prior on ω^2 with prior mean 0.25, 1 000 draws plus 100 for warming the chain. Posterior draws were then used to

Table 2: Poverty rate and intensity in East Germany 2002-2006

	Poverty rate			Poverty intensity		
	Total	Chronic	Transient	Total	Chronic	Transient
Child disposable	0.218	0.071	0.147	0.063	0.012	0.051
	(0.012)	*(0.013)*	*(0.011)*	*(0.005)*	*(0.003)*	*(0.004)*
Child market	0.496	0.208	0.289	0.372	0.116	0.255
	(0.030)	*(0.033)*	*(0.015)*	*(0.027)*	*(0.022)*	*(0.012)*
Adult disposable	0.185	0.052	0.134	0.050	0.008	0.042
	(0.007)	*(0.006)*	*(0.005)*	*(0.003)*	*(0.001)*	*(0.002)*
Adult market	0.649	0.355	0.294	0.557	0.253	0.304
	(0.022)	*(0.032)*	*(0.012)*	*(0.023)*	*(0.026)*	*(0.008)*

Standard errors are given between parentheses. Market income represent total income before taxes and redistribution. Disposable income includes taxes and redistribution and in particular family allowances. Child corresponds to the population between 1 and 18 years. Adult are over 18 years and have no child.

simulate incomes and poverty indices with results reported in Table 2. Before taxes and transfers, there is much more poverty among adults as if poor adults had decided not to have children. Poverty among adults is mostly chronic when it is mainly transitory among children. Poverty intensity is also stronger among adults while being mostly transient. When taxes and transfers are introduced, total poverty is much reduced, but the reduction is more important among adults than among children. With transfers, child and adult poverty become mainly transient while chronic poverty intensity is reduced to very low levels. We have thus a contrasted impact of social transfers on the dynamic of poverty in East Germany for that period. As underlined in Fourrier-Nicolaï and Lubrano (2020), the major changes introduced after the Hartz plan reforms in the German redistributive system after 2006 contributed a lot to reduce child poverty.

5.3 Poverty dynamics using pseudo panels

Poverty dynamics can be analysed using a bivariate dynamic probit model which explains between two periods the transition between two states, poor and non-poor. Stayers or chronic poverty is being poor both at $t-1$ and t. Poverty entry is not being poor at $t-1$ and entering poverty at t, transitory poverty is being poor at $t-1$ and getting out of poverty at t. Cappellari and Jenkins (2004) considers three equations for explaining the poverty status dynamics along with a correction for attrition explained in a third equation as individuals might not be missing at random. This model requires consecutive observations to build up data pairs for dynamic analysis, which are not always available, especially in developing countries. The purpose of Sadeq and Lubrano (2018) was to use an adapted version of this model to measure the impact of the Wall built on the West Bank after 2002 on poverty dynamics in occupied territories. In two repeated cross-sectional waves of 2004 and 2011 from *Palestinian Expenditure and Consumption Survey*, a variable indicated if a household was impacted or not by the Wall. Sadeq and Lubrano (2018) considered a pseudo panel as an incomplete data problem. Inside the loop of a Gibbs sampler, they explain the income-to-needs ratio (negative for being under poverty) using time invariant data for 2004 and 2011 and the grouping techniques of Deaton (1985) and Verbeek and Vella (2005) to generate the missing values and recover information on ρ, the correlation parameter between the two periods error terms. Then they use both observed and latent variables to explain the income-to-needs ratio for 2011, this time conditionally on being poor in 2004 and being affected or not by the Wall. We have thus two ways of measuring poverty dynamics and the final effect of the wall on poverty dynamics is determined by the difference between a marginal probability and a conditional probability taking into account the effect of the Wall. The paper is rather technical and its equations will not be fully detailed here.

Taking into account the Wall has a large effect on poverty dynamics. For those who were already poor in period 1, the wall increases their probability of staying poor by 58 percentage points. For those who were not in poverty, the probability of entering into poverty during the second period is increased by 18 percentages points.

We have reproduced in Figure 3 the posterior densities of these probabilities, using plain lines. We compare these probabilities to those obtained under a non-informative prior on ρ, using dashed lines. With a non-informative prior on ρ, the differential in probability of poverty entry is slightly increased while the differential in poverty persistence is slightly decreased. But these differences are mild, even if the prior information had a sizable influence on the posterior density of ρ.

6 Conclusion and further reading

The reader might understand that we made a restricted presentation. We assumed most of the time that the income distribution could be represented by a mixture of lognormals. The Double Pareto lognormal distribution of Reed and Jorgensen (2004) could be an alternative. For Bayesian inference using a Gibbs sampler, see Ramirez-Cobo et al. (2010).

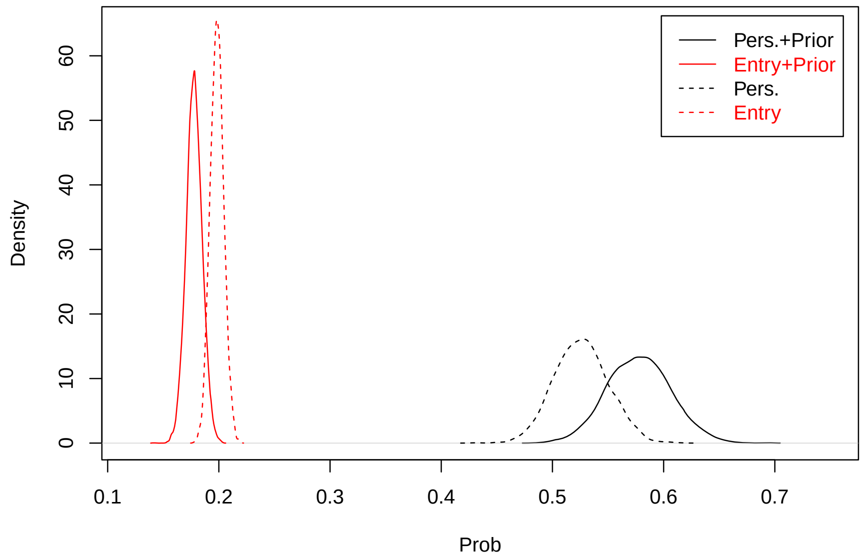


Figure 3: Posterior density of poverty persistence
and poverty entry differentials due to the Wall

We assumed that individual survey data were available. In many cases only group data are available. Groups can correspond to fixed bounds reporting the number of households inside each cell. For instance the *American Community Survey* provides income data at the level of School Districts in the form of ten income classes. Groups can have variable bounds, each group containing the same proportion of individuals. This is convenient for reporting income shares as does the World Inequality data base. To each case corresponds a specific statistical problem as surveyed in Eckernkemper and Gribisch (2020). Chotikapanich and Griffiths (2000), Griffiths et al. (2005), Chotikapanich and Griffiths (2005), Kakamu (2016), Kakamu and Nishino (2019) contributed also to this field.

For many authors, the spatial dimension has to be taken into account for measuring poverty. Haughton and Phong (2003) estimate poverty rates in Vietnam provinces, considering poverty as a binomial process (poor and non-poor) within clusters represented by administrative districts that are then aggregated. Wieczorek and Hawala (2011) study spatial poverty in the US at the county level. County poverty rates μ_i are then explained in a logit model. Their main objective is to to predict poverty rates, taking into account county size and sampling design. Nawawi et al. (2020) use a Poisson

Log-Linear Leroux Conditional Autoregressive model with different neighbourhood matrices for explaining poverty rates in 66 districts of Kelantan, Malaisia, in 2010 by various socio-economic indicators.

References

Atkinson, A. B. and Bourguignon, F. (2001). Poverty and inclusion from a world perspective. In Stiglitz, J. E. and Muet, P.-A., editors, *Governance, Equity and Global Markets*. Oxford University Press.

Biewen, M. and Jenkins, S. P. (2006). Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case. *Oxford Bulletin of Economics and Statistics*, 68(3):371–383.

Brent, R. P. (1971). An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, 14(4):422–425.

Cappellari, L. and Jenkins, S. P. (2004). Modelling low income transitions. *Journal of Applied Econometrics*, 19(5):593–610.

Chib, S. (1996). Inference in panel data models via Gibbs sampling. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data*, volume 33 of *Advanced Studies In Theoretical and Applied Econometrics*, chapter 24, pages 639–651. Kluwer Academic Publishers, Dordrecht, The Netherlands, second edition.

Chib, S. and Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, 9:17–26.

Chotikapanich, D. (1993). A comparison of alternative functional forms for the Lorenz curve. *Economics Letters*, 41(2):129 – 138.

Chotikapanich, D. and Griffiths, W. E. (2000). Posterior distributions for the Gini coefficient using grouped data. *Australian and New Zealand Journal of Statistics*, 42:383–392.

Chotikapanich, D. and Griffiths, W. E. (2005). Averaging Lorenz curves. *Journal of Economic Inequality (2005)*, 3:1–19.

Corak, M., Fertig, M., and Tamm, M. (2008). A portrait of child poverty in Germany. *Review of Income and Wealth*, 54(4):547–571.

Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics*, 150(1):30 – 40.

Davidson, R. and Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, 68:1435–1464.

Davidson, R. and Duclos, J.-Y. (2013). Testing for restricted stochastic dominance. *Econometric Reviews*, 32(1):84–125.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30:109–126.

Eckernkemper, T. and Gribisch, B. (2020). Classical and Bayesian inference for income distributions using grouped data. *Oxford Bulletin of Economics and Statistics*.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52:761–765.

Fourrier-Nicolaï, E. and Lubrano, M. (2020). Bayesian inference for TIP curves: an application to child poverty in Germany. *The Journal of Economic Inequality*, 18:91–111.

Fourrier-Nicolaï, E. and Lubrano, M. (2021). Bayesian inference for parametric growth incidence curves. In Bandyopadhyay, S., editor, *Research on Economic Inequality*, volume 29. Emerald Group Publishing Limited.

Gastwirth, J. (1971). A general definition of the Lorenz curve. *Econometrica*, 39(6):1037–1039.

Griffiths, W. E., Chotikapanich, D., and Rao, D. S. P. (2005). Averaging income distributions. *Bulletin of Economic Research*, 57(4):347–367.

Gunawan, D., Panagiotelis, A., Griffiths, W., and Chotikapanich, D. (2020). Bayesian weighted inference from surveys. *Australian and New Zealand Journal of Statistics*, 62(1):71–93.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.

Hasegawa, H. and Ueda, K. (2007). Measuring chronic and transient components of poverty: a Bayesian approach. *Empirical Economics*, 33:469–490.

Haughton, D. and Phong, N. (2003). Bayesian analysis of poverty rates: The case of Vietnamese provinces. *Journal of Modern Applied Statistical Methods*, 2(1):189–194.

Jawitz, J. W. (2004). Moments of truncated continuous univariate distributions. *Advances in Water Resources*, 27(3):269–281.

Jenkins, S. P. and Lambert, P. J. (1997). Three 'I's of poverty curves, with an analysis of UK poverty trends. *Oxford Economic Papers*, 49(3):317–327.

Kakamu, K. (2016). Simulation studies comparing Dagum and Singh–Maddala income distributions. *Computational Economics*, 48:593–605.

Kakamu, K. and Nishino, H. (2019). Bayesian estimation of beta-type distribution parameters based on grouped data. *Computational Economics*, 54:625–645.

Kakwani, N. (1980). On a class of poverty measures. *Econometrica*, 48(2):437–446.

Kakwani, N. and Podder, N. (1973). On the estimation of Lorenz curves from grouped observations. *International Economic Review*, 14:278–292.

Lander, D., Gunawan, D., Griffiths, W., and Chotikapanich, D. (2020). Bayesian assessment of Lorenz and stochastic dominance. *Canadian Journal of Economics / Revue canadienne d'économique 2020*, 53(2):767–799.

Lindley, D. V. (1971). *Bayesian Statistics, A Review.* CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.

Lubrano, M. and Ndoye, A. A. J. (2016). Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. *Computational Statistics and Data Analysis*, 100:830 – 846.

Muller, C. (2001). The properties of the Watts poverty index under lognormality. *Economics Bulletin*, 9(1):1–9.

Nawawi, S. A., Busu, I., Fauzi, N., Amin, M. F. M., and Yusof, N. R. N. (2020). Relative risk for poverty in Kelantan: A Bayesian approach. *IOP Conference Series: Earth and Environmental Science*, 549:012079.

Ndoye, A. and Lubrano, M. (2014). Tournaments and superstar models: A mixture of two Pareto distributions. In *Economic Well-Being and Inequality: Papers from the Fifth ECINEQ Meeting (Research on Economic Inequality, Vol. 22)*, pages 449–479. Emerald Group Publishing Limited.

Ramirez-Cobo, P., Lillo, R. E., Wilson, S., and Wiper, M. P. (2010). Bayesian inference for double Pareto lognormal queues. *The Annals of Applied Statistics*, 4(3):1533–1557.

Ravallion, M. (1988). Expected poverty under risk-induced welfare variability. *The Economic Journal*, 98(393):1171–1182.

Ravallion, M. and Chen, S. (2003). Measuring pro-poor growth. *Economics Letters*, 78:93–99.

Ravallion, M., Chen, S., and Sangraula, P. (2009). Dollar a day revisited. *World Bank Economic Review*, 23:163–184.

Reed, W. J. and Jorgensen, M. (2004). The double Pareto-lognormal distribution: A new parametric model for size distributions. *Communications in Statistics - Theory and Methods*, 33(8):1733–1753.

Sadeq, T. and Lubrano, M. (2018). The wall's impact in the occupied West Bank: A Bayesian approach to poverty dynamics using repeated cross-sections. *Econometrics*, 6(2):9.

Shorrocks, A. F. (1995). Revisitng the Sen poverty index. *Econometrica*, 63(5):1225–1230.

Verbeek, M. and Vella, F. (2005). Estimating dynamic models from repeated cross-sections. *Journal of Econometrics*, 127:83–102.

Villasenor, J. A. and Arnold, B. C. (1989). Elliptical Lorenz curves. *Journal of Econometrics*, 40:327–338.

Watts, H. (1968). An economic definition of poverty. In Moynihan, D., editor, *On Understanding Poverty*. Basic Books, New-York.

Wieczorek, J. and Hawala, S. (2011). A Bayesian zero-one inflated beta model for estimating poverty. Working paper, US Bureau of Census.

Xun, Z. and Lubrano, M. (2018). A Bayesian measure of poverty in the developing world. *Review of Income and Wealth*, 64(3):649–678.

Zheng, B. (1997). Aggregate poverty measures. *Journal of Economic Surveys*, 11(2):123–162.