Data driven models for wealth inequality

Convergence of empirical Lorenz curves

Daniel Watson
Supervisor: Enrico Scalas

A dissertation presented for the degree of Masters of Data Science

School of Mathematical and Physical Sciences
University of Sussex
August 2020

Contents

1 Abstract **2**

2 Definition of Lorenz curves **2**
2.1 Infinite mean Lorenz curves . . . 3

3 The ethics of inequality **3**

4 Capital in the Twenty-First Century **3**
4.1 Relevance to Lorenz curves . . . 3
4.2 The long term trends of inequality . . . 4
4.3 Bringing inequality back to the center of economics analysis . . . 4

5 Theoretical and empirical Lorenz curves **5**
5.1 Additional references . . . 5
5.2 Formal definition . . . 5
5.3 Empirical Lorenz curves . . . 6
5.4 Properties . . . 6

6 Gini coefficient **7**

7 Monte Carlo simulation to illustrate convergence **8**

8 Generalized Lorenz curves **10**

9 Conjecture **10**
9.1 Definitions and preliminaries . . . 10
9.2 Convergence under J_1 topology . . . 13

10 Other inequality measures **15**
10.1 Pietra index . . . 15
10.2 Atkinson index . . . 15
10.3 Generalised Entropy family . . . 16

11 Inequality measurement requirements **16**

12 Pigou-Dalton Monte Carlo simulations **17**
12.1 Flat taxation . . . 17
12.2 Robin-hood style (quantile) taxation . . . 19
12.3 Emulating the UK tax system . . . 21

13 Generating Mittag-Leffler data and fitting to a power-law distribution **24**

14 Empirical work with the British Household Panel Survey **27**
14.1 Fitting to a log-normal distribution . . . 28
14.2 A Bayesian implementation for the log-normal distribution . . . 30
14.3 Bootstrapping . . . 32
14.4 A Bayesian implementation for the Gamma distribution . . . 34

15 Empirical work with Office of National Statistics data **35**
15.1 Lorenz curves and data manipulation . . . 35
15.2 First order stochastic dominance . . . 37
15.3 Stochastic dominance t-test . . . 39
15.4 Second order stochastic dominance . . . 41
15.5 Transition matrices . . . 43
15.6 Taxation revisited . . . 50

16 Conclusion **55**

17 References **55**

1 Abstract

In 1905, Max O. Lorenz conceptualised the "Lorenz curve" which depicts the level of inequality as a graphical representation (Lorenz, 1905). Throughout the dissertation we use Gastwirth's definition of the Lorenz curve which applies to a probability distribution (Gastwirth, 1971). It applies to both a theoretical and empirical distribution, where we aim to prove that Lorenz curves of the latter converge to the former when the distribution has infinite mean. Data analysis relating to empirical Lorenz curves is also carried out in the dissertation, utilising data from two sources regarding inequality in the UK. Examples of explored topics include stochastic dominance, Markov models and Monte Carlo simulations.

2 Definition of Lorenz curves

Lorenz curves have numerous applications in different areas, with an example being inequality among plant sizes and their fertilities in the field of ecology (Damgaard and Weiner, 2000). However, Lorenz curves are traditionally used to highlight wealth or income inequality in societies, with the cumulative share of the populace lying on the x-axis and the cumulative share of the income or wealth they hold on the y-axis. If concerned with income inequality, it can be said that for the bottom $p_1\%$ individuals, what percentage $p_2\%$ of the total income they earn. Figure one represents an example. Perfect equality would be represented by a straight line. Realistically, the curve is defined below this line which is of course due to the fact that inequality persists in global societies.

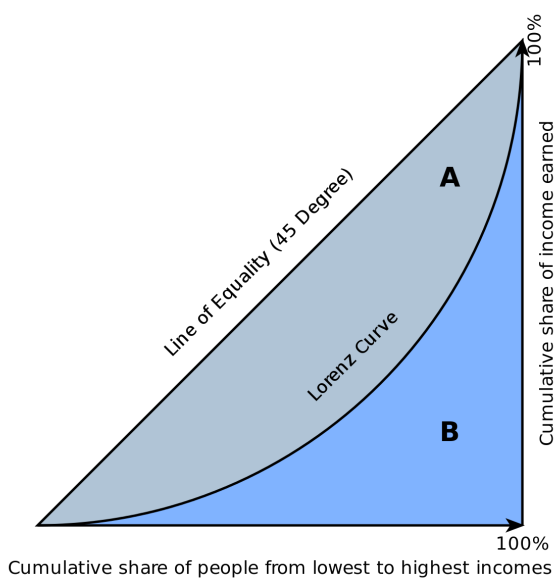


Figure 1: A basic Lorenz curve [27]

2.1 Infinite mean Lorenz curves

A Lorenz curve with infinite mean would depict a situation of extreme inequality whereby one or a small number of individuals own all of the wealth available in the concerned society. The distribution we are most concerned with is the Mittag-Leffler distribution produced by the one-parameter Mittag-Leffler function which will be defined later. This is a heavy-tailed infinite mean distribution with large outliers and so depicts the situation of extreme inequality.

Charles Goldie's 1977 paper "Convergence Theorems for Empirical Lorenz Curves and Their Inverses" is the most relevant source of information for the conjecture. This is because the conjecture can be seen as a special case of Goldie's previously established result whereby he proves uniform convergence of the sequence of empirical Lorenz curves to the theoretical Lorenz curve for finite mean. He proves a similar for their inverses which known as concentration curves.

3 The ethics of inequality

Before analysing the long term trends of inequality and the reasons for the shift over time, (with particular relevance to the last 300 years or so) it is important to speak about why we are interested in inequality itself. While inequality seems to inherently be a constant force in human society, it is of course often seen as unacceptable due the power and status that wealth brings to the individual. This, culminated with the fact that vast wealth is usually only attainable through inheritance naturally causes conflict when discussing inequality. One could argue that all wealth attained is a result of entrepreneurship and hence inequality is justified. While this is true to some extent and that vast earned wealth through this process should be respected, this wealth is commonly secured in part through the exploitation of natural resources such as land or the profiteering off of unethical practices involving human labour (Basu 2018). Piketty declares in his book that "every fortune, although historically partially justified, is potentially excessive" - a quote that is very applicable to the current global economical climate in which a select few individuals hold enormous fortunes. The 21st century so far seems destined for a society comprised of ultra-wealthy influential figures that transcend levels of traditional wealth.

4 Capital in the Twenty-First Century

4.1 Relevance to Lorenz curves

In 2013 the French economist Thomas Piketty published his *magnum opus*, "Capital in the Twenty-First Century". This book provides the motivation for the dissertation due

to the fact that it re-ignited discussion regarding inequality after its release. The central theme presented is that of wealth and income inequality whereby Piketty goes into meticulous detail surrounding the subject, with a particular focus on the how concentration of wealth arises and the forces that drive it. He argues that the unequal distribution of wealth is in large part the result of the rate of return on capital being higher than that of the rate of growth. Piketty makes heavy use of very long term data, offering a clear view of the trends over time which also allow him to make conclusions with substantial evidence. The reason for Piketty's book being a major source of reference for this dissertation is in the fact that his data analysis usually involves quantiles, which can be used to define Lorenz curves and so is directly relevant. The purpose of the heavy use of quantiles is that it provides independence of opinion - with the most common classification of income or wealth groups being the working, middle and upper class. This helps retain the universality of comparison across all global societies and as Piketty states, allows him to make comparisons without undermining the inherent complexity of societies.

4.2 The long term trends of inequality

Throughout the book, it becomes clear that the trends regarding inequality have fluctuated greatly, with particular volatility throughout the 20th century. From antiquity until the industrial revolution, global growth was low both in the demographic and economic sense, with economic growth perpetually stuck at around 0.1 - 0.5%. This was due to a large variety of factors with the most obvious being the low level of productivity caused by the slow progress of technological innovation. During the industrial revolution however, economic growth steadily began to rise and has more-or-less maintained this level until present day, at around 3.5%. This of course caused the rise in living standards, life expectancy and consumer purchasing power but also helped aid the rate of return on capital. This issue arises when it becomes apparent that the rate of return on capital has consistently been greater than the rate of growth across the globe, resulting in the concentration of wealth as described before.

4.3 Bringing inequality back to the center of economics analysis

Piketty also brings to light the negligence of 20th century economists regarding the distribution of wealth and incomes, whereby it has been disregarded. This contrasted with the 19th century where economists such as David Ricardo and Karl Marx raised their concerns with their iconic works including the "Principles of Political Economy and Taxation" by Ricardo in 1817 and "Capital" by Marx in 1867. These works highlighted the issues of infinite accumulation through the increasing share of income and output claimed by the rich and the reasons for these predictions. This of course would result in large inequalities in the respective society over time and so is of central interest to the study of inequality. However, both did not have access to the statistical data they needed to

reinforce their predictions due to a lack of public records. By the end of the 19th century it also came to light that Marx and Ricardo's forecasts were off the mark, with consumer purchasing power increasingly rapidly and the standard of living of the general populace rising swiftly. In Marx's case, Piketty attributed this to his oversight of the process of steadily increasing productivity and sustainable technological progress.

Despite this, inequality itself did persist throughout the dawn of the 20th century up until the eve of World War I. The following shocks to global society during the onset of World War I and the continuation into World War II saw a fall in inequality which remained at a stable level until around 1970. This prompted 20th century economists like Simon Kuznets to propose that inequality decreases as capital societies become more developed, following the trend of the "Kuznets curve". Inequality however began to increase again after 1970 and has once again began to reach levels apparent in the pre-war era. Piketty suggests ways to combat these growing concerns of large levels of inequality in the 21st century in his book, with an example being a global progressive tax on capital to help prevent further divergence between the rate of return on capital and the rate of economic growth.

5 Theoretical and empirical Lorenz curves

5.1 Additional references

The other extremely important reference we make use of in addition to Piketty's book and Goldie's paper is Michel Lubrano's lecture notes on "The econometrics of inequality and poverty measurement" at the University of Marseilles. This provides significant background technical knowledge regarding Lorenz curves and related econometric concepts and provides inspiration for many of the later sections.

5.2 Formal definition

We now give the formal definition of a Lorenz curve, with some prerequisites: Start from a distribution $f(x)$, having mean:

$$\mu = \int_0^\infty x f(x) dx = \int_0^\infty x dF(x) = \int_0^1 q(p) dp \tag{1}$$

If we apply a change of variable $x = F^{-1}(p) = q(p) = \inf\{x : F(x) \geq p\}$ and $p = F(x)$, differentiating gives us $dp = f(x)dx$ which means that:

$$\mu = \int_0^1 F^{-1}(p) dp \tag{2}$$

where p corresponds to a proportion of the population and $q(p)$ is the living standard below which we find this proportion - a quantile. It then follows that:

$$p = F(z) = \int_0^z f(t)dt \tag{3}$$

giving:

$$L(p) = \frac{1}{\mu}\int_0^z tf(t)dt \tag{4}$$

We then get:

$$L(p) = \frac{1}{\mu}\int_0^p q(t) = \frac{1}{\mu}\int_0^p F^{-1}(t)dt \tag{5}$$

Using (1) we can also rewrite this expression as:

$$L(p) = \frac{\int_0^p q(t)dt}{\int_0^1 q(t)dt} \tag{6}$$

which provides the most intuitive take on the Lorenz curve, with it being the ratio of total income or wealth below the proportion p of the population against the total wealth or income of the population as a whole.

5.3 Empirical Lorenz curves

Now that the theoretical Lorenz curve is defined, we can move onto the definition of the empirical Lorenz curve. We use the definition presented in (Gastwirth, 1972) which is also utilised in Goldie's paper. The empirical Lorenz curve L_n is constructed from the empirical distribution determined by the sample $X_1, \ldots, X_n$. More rigorously, if we take the order statistics of $X_1, \ldots, X_n$ to get $X_{1n} \leq X_{2n} \leq \cdots \leq X_{nn}$ and define $S_{jn} = \sum_{i=1}^{j} X_{in}$ for $j = 1, \ldots, n$ then L_n is the polygonal line joining the points $(j/n, S_{jn}/S_{nn})$. This will be expanded upon later when considering the conjecture.

5.4 Properties

It is useful to present certain mathematical properties of Lorenz curves as to provide some intuition:

- p and $L(p)$ both lie on [0,1], hence the Lorenz curve is bounded on $[0, 1]^2$.
- If the mean of the distribution used is infinite, the Lorenz curve is not defined.
- Lorenz curves are invariant with positive scaling.
- $\dfrac{dL(p)}{dp} = \dfrac{q(p)}{\mu} = \dfrac{x}{\mu}$. This means that at the mean when $q(p) = \mu$, $\dfrac{dL(p)}{dp} = 1$ and so the slope of the Lorenz curve at the mean is exactly one.

6 Gini coefficient

Now that the Lorenz curve is rigorously defined, we can provide a definition of the Gini coefficient, the most widely used inequality metric. One way that the Gini coefficient is defined is as the ratio of the area between the perfect line of equality and the Lorenz curve in question over the total area below the perfect line of equality. This would correspond to A/A+B when looking Figure one. The line of perfect equality is of course given by a straight line meaning that $L(p) = p$. Hence $p - L(p)$ describes this difference and so:

$$G = 2\int_0^1 (p - L(p))dp = 2\int_0^1 pdp - 2\int_0^1 L(p)dp = 1 - 2\int_0^1 L(p)dp \tag{7}$$

The Gini coefficient is traditionally seen as the index that best encapsulates inequality in societies. This is in part due to the fact that the coefficient measures the difference in incomes between every pairwise combination in the sample which is then normalised by the mean, therefore assessing every individuals relative wealth. This gives an alternative definition where if we have a distribution of wealth defined by $f(x)$:

$$G = \frac{1}{2\mu}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(y)|x - y|dxdy \tag{8}$$

where μ is the mean of the distribution. A result of zero would result in the Lorenz curve in question being equal to the line of perfect equality hence indicating that there is perfect equality. As a result, a coefficient of one indicates perfect inequality. Therefore in the case of the conjecture, we would also see a convergence of the Gini to one. There are more definitions of the Gini coefficient including as a covariance but these are the two most relevant to Lorenz curves, with the former being of particular interest.

However, of course the Gini coefficient has its limitations which begin with the fact that it solely takes into account the individuals current wealth, ignoring their potential future wealth. Summing inequality into one figure of course means that certain information is not encapsulated by the coefficient and so two different distributions can produce the same index with one clearly being more unequal than the other upon further inspection (Lamb 2012).

7 Monte Carlo simulation to illustrate convergence

In this section we perform a Monte Carlo simulation to display the convergence of the sequence of empirical Lorenz curves to that of perfect inequality when considering a distribution with infinite mean. Our distribution of choice is the Mittag-Leffler distribution which is generated from the one parameter Mittag-Leffler function given below in (10). Using equation (20) from (Fulger, Scalas and Germano, 2008) we can generate Mittag-Leffler random variables which are given by:

$$\tau_\beta = -\gamma_t \log(u) \left(\frac{\sin(\beta\pi)}{\tan(\beta\pi v)} - \cos(\beta\pi) \right)^{1/\beta} \tag{9}$$

where u, v are independent standard uniform random variates, γ_t is the scale parameter and β is defined by the Mittag-Leffler function which is:

$$E_\beta(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\beta k + 1)} \tag{10}$$

We keep the value of gamma to be constant at $\gamma_t = 1/12$ when using the above formula. Figure two below represents the simulation, whereby choosing $\beta = 0.6$ produces very high levels of inequality. $\beta = 0.95$ produces a much more equal curve but due to the fact that the distribution has infinite mean, this will eventually reach a case of perfect inequality with $n \to \infty$, looking more like the situation at $\beta = 0.6$.

The table below also shows the Gini coefficient for both the empirical and theoretical distributions for increasing sample size. As expected, the tending of the coefficient to a value of one corresponds to the convergence of the curves. The leftmost figure in Figure two with $\beta = 0.6$ hence represents our conjecture directly.

Below is table of the Gini indices corresponding to their respective sample sizes, showing an increase in inequality as sample size increases.

Sample size	Gini coefficient
100	0.934
1000	0.997
10000	0.998

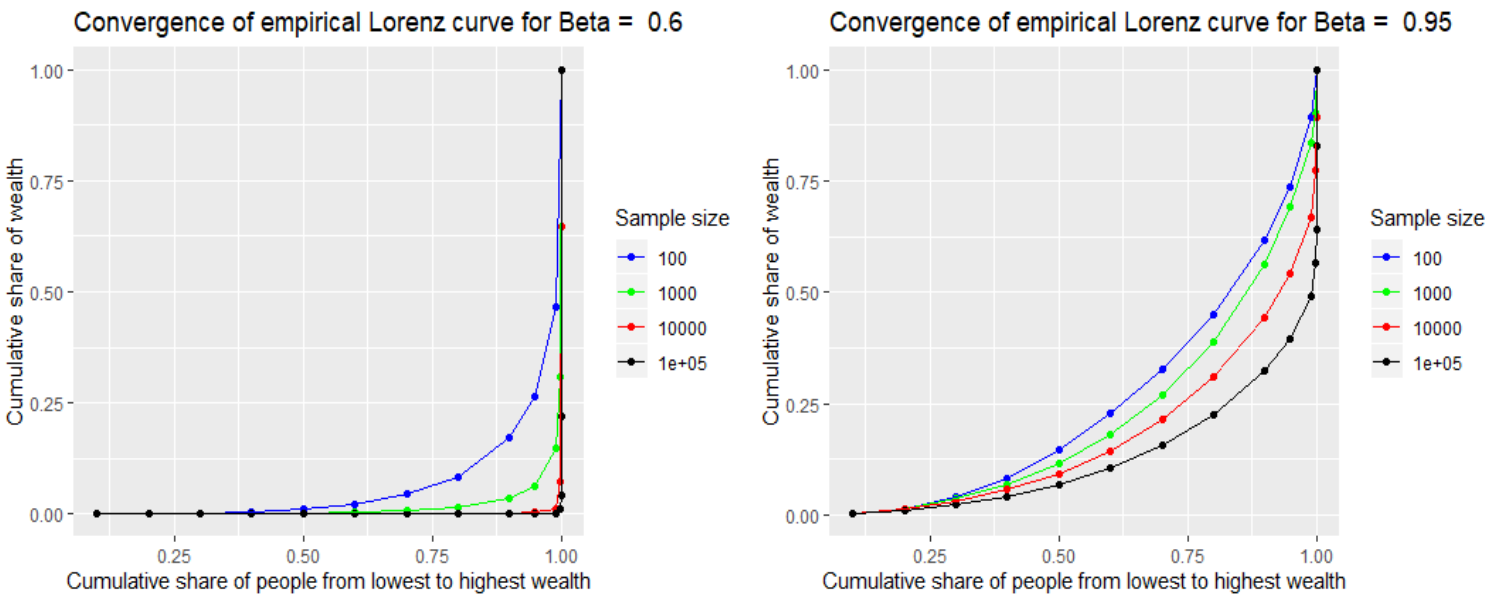


Figure 2: Convergence of empirical Lorenz curve as sample size increases with beta = 0.95 and beta = 0.6

In addition we can vary β and find the Gini index that corresponds to the individual β value. This was repeated 1000 times with 1 million samples at a time, producing the plot below:

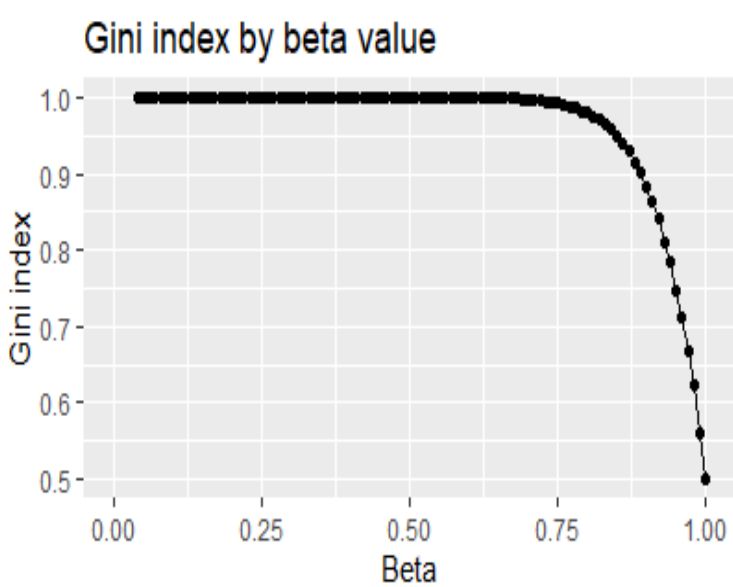


Figure 3: Relationship between β values for ML-random variables and the Gini index produced

Clearly as the value of β rises, the Gini index begins to fall, with perfect inequality persisting for β values at and below around 0.75. After this we see a sharp fall in the level of inequality praesent in the distribution, with $\beta = 1$ having the lowest inequality

with an index of 0.5. Note that this plot would be different for various other values of n, which would most likely have the same shape but a different base value of the Gini.

8 Generalized Lorenz curves

The Lorenz curve is scale invariant meaning that if our distribution $f(x)$ becomes $cf(x)$ whereby the incomes or wealth of every individual in the society is multiplied by some constant c then this has no effect on the Lorenz curve. Hence every individual could become richer, enjoying higher living standards yet the new Lorenz curve produced would not reflect this.
This gives the motivation for the generalized Lorenz curve which is defined simply as the mean multiplied by the original Lorenz curve, therefore removing the normalisation provided by the mean (Shorrocks, 1983). The definition is:

$$GL(p) = \mu L(p) = \int_0^p F^{-1}(t)dt \tag{11}$$

This means that more emphasis is also placed on the mean standard as well as general inequality whereas the original Lorenz curve is not concerned with the value of the mean. In later sections this is explored further, in particular the relationship with stochastic dominance.

9 Conjecture

This section provides the proof of the conjecture, with the final result being that L_n converges under the J_1 topology to L i.e $d_{J_1}(L_n, L) \overset{n\to\infty}{\to} 0$.

Conjecture 1. *If we have a distribution function with $\mu = \infty$, $L_n \overset{n\to\infty}{\to} L$ where L represents the Lorenz curve in a situation of perfect inequality which is given by:*

$$L(p) = \begin{cases} 0, & p \in [0,1) \\ 1, & p = 1 \end{cases}$$

9.1 Definitions and preliminaries

We can define the Lorenz curve in terms of the cumulative distribution function F of the probability distribution function on which L is defined, as well as $G(t)$ as defined below:

$$G(t) = \frac{1}{\mu} \int_{[0,t]} xF(dx) \tag{12}$$

From Goldie's paper we find that if the probability distribution has an atom at $t > 0$ which are the points of the sample space with positive probability, then both F and G are discontinuous there and so the Lorenz curve includes the closed straight-line segment

joining $(F(t), G(t-))$ and $(F(t+), G(t))$. It is also assumed that G is right-continuous and F is left-continuous. At other points $t \in [0, \infty]$ both F and G will be continuous and so we let the graph include the point $(F(t), G(t))$. As a reminder, the Lorenz curve can be given in the form:

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(x)dx, 0 \leq p \leq 1. \tag{13}$$

The derivation of this was given in Section 5. The empirical Lorenz curve L_n is constructed from the empirical distribution determined by the sample $X_1, \ldots, X_n$. More rigorously, if we take the order statistics of $X_1, \ldots, X_n$ to get $X_{1n} \leq X_{2n} \leq \cdots \leq X_{nn}$ and define $S_{jn} = \sum_{i=1}^{j} X_{in}$ for $j = 1, \ldots, n$ then L_n is the polygonal line joining the points $(j/n, S_{jn}/S_{nn})$. This results in $G_n = S_{nn}^{-1} \sum_{i=1}^{n} X_{i\chi[X_i=t]}, 0 \leq t \leq \infty$ where $\chi_{[\cdot]}$ is the function taking the value one if its argument is true or zero if false.

Lemma 2. ***Chung's Lemma (Chung 1968, p124)*** *Let f_n, f be real-valued non-decreasing right-continuous functions defined on a closed interval $[a, b]$ where $-\infty \leq a < b \leq \infty$ and let $T \subseteq [a, b]$ be dense in $[a, b]$ and contain a and b. Suppose that $f_n(t) \to f(t)$ for all $t \in T$ and that for each t in the set T_f of discontinuities of f we have $f_n(t) - f_n(t-) \to f(t) - f(t-)$. Then f_n converges to f uniformly on $[a, b]$.*

Because we have infinite mean, $G(t) = 0 \, \forall \, t \in [0, \infty)$ and $G(\infty) = 1$, with the following Lemma proving convergence of G_n.

Lemma 3. $G_n \overset{n \to \infty}{\to} G$ *uniformly on* $[0, \infty)$.

Proof. In Lemma 2 of his paper, Goldie uses the strong law on the i.i.d sequences $\{X_{i\chi[X_i=t]}\}$ and $\{X_i\}$ whose corresponding partial sums is equal to $G_n(t) - G_n(t-)$ for $t \in T_G$, the set of discontinuities of G. This gives:

$$G_n(t) - G_n(t-) \overset{a.s}{\to} t[F(t+) - F(t)]/\mu = G(t) - G(t-) \tag{14}$$

through the application of the Glivenko-Cantelli theorem. The strong law can be used in a similar way for each rational $t \geq 0$ on the i.i.d sequences $\{X_{i\chi[X_i \leq t]}\}$ and $\{X_i\}$ to give $G_n(t) \overset{a.s}{\to} G(t)$. The result follows through the application of Chung's Lemma. □

Below we include a proof of the Glivenko-Cantelli theorem but for the inverse cumulative distribution function and its empirical counterpart, where is of course given by:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[X_i, \infty)}(x) = \frac{1}{n} |\{1 \leq i \leq n | X_i \leq x\}| \tag{15}$$

Additionally, we can define $F^{-1}(p) = \sup\{t : F(t) \leq p\}$

Lemma 4. $\|F_n^{-1} \to F^{-1}\|_\infty \overset{n \to \infty}{\to} 0$*:*

Proof. If we have $-\infty = x_0 < x_1 < \cdots < x_{m-1} < x_m = \infty$ such that $F^{-1}(x_{j-1}) - F^{-1}(x_j) = 1/m$ for $j = 1, \ldots, m$. Then $\forall x \in \mathbb{R}$ there exists $j \in \{1, \ldots, m\}$ such that $x \in [x_{j-1}, x_j]$. Then:

$$F_n^{-1}(x) - F^{-1}(x) \geq F_n^{-1}(x_j) - F^{-1}(x_{j-1}) = F_n^{-1}(x_j) - 1/m - F^{-1}(x_j) \tag{16}$$

$$F_n^{-1}(x) - F^{-1}(x) \leq F_n^{-1}(x_{j-1}) - F^{-1}(x_j) = F_n^{-1}(x_{j-1}) + 1/m - F^{-1}(x_{j-1}) \tag{17}$$

Therefore, almost surely, $\|F_n^{-1}(x) - F^{-1}(x)\|_\infty = \sup_{x\in\mathbb{R}} |F_n^{-1}(x) - F^{-1}(x)|$
$\leq \max_{j\in\{1,\ldots,m\}} |F_n^{-1}(x_{j-1}) - F^{-1}(x_{j-1})| + 1/m$.

$\max_{j\in\{1,\ldots,m\}} |F_n^{-1}(x_{j-1}) - F^{-1}(x_{j-1})| \to 0$ almost surely due to the strong law of large numbers and so for any m we can find an integer N such that for all $n \geq N$ we have:

$$\|F_n^{-1}(x) - F^{-1}(x)\|_\infty \leq 1/m \tag{18}$$

giving uniform convergence where $\|f\|_\infty = \sup\{|f(x)| : x \in S\}$ when f is defined on a set S. In the case of $F(x), S = [0, \infty]$ □

The following Lemma is copied from Goldie's paper (p 769) for convenience. Let A be the probability-one event that $F_n \overset{u}{\to} F$ and that $G_n \overset{u}{\to} G$ where u denotes uniform convergence.

Lemma 5. *On A, for each $p \in (0, 1)$ the accumulation points of the sequence $\{F^{-1}(p)\}_{n=1}^{\infty}$ can only be $F^{-1}(p-)$ and $F^{-1}(p)$, and if F^{-1} is continuous at p then $F_n^{-1}(p) \to F^{-1}(p)$.*

Proof. For any $t > F^{-1}(p)$ we have $F_n(t) \to F(t) > p$ so that $F_n^{-1}(p) \leq t$ for large n. Likewise, for any $t < F^{-1}(p-), F_n(t) \to F(t) < p$ so that $F_n^{-1}(p) \geq t$ for large n. Therefore $F^{-1}(p-) \leq \lim_n \inf F_n^{-1}(p) \leq \lim_n \sup F_n^{-1}(p) \leq F^{-1}(p)$. Values of F_n^{-1}, being in S, cannot lie in the open interval $(F^{-1}(p-), F^{-1}(p))$. The first statement of the lemma follows, and the second is an immediate consequence where $S \subseteq (0, \infty)$ is the support of the probability distribution. □

Lemma 6. *$L_n \overset{n\to\infty}{\to} L$ pointwise.*

Proof. If we return to Lemma 3 we can see that the set of discontinuities of G denoted T_G would contain the values of t which are atoms of the probability distribution. Conversely each rational $t \geq 0$ where $G_n \overset{a.s}{\to} G$ are not atoms, implying continuity of G at that point. The set T_G of discontinuities of G would only be the point where $t = \infty$ meaning that $G_n(\infty) - G_n(\infty-) \overset{a.s}{\to} G(\infty) - G(\infty-)$. Clearly $G(\infty) = 1$ from the definition of G, and due to the fact that $G(t) = 0$ for any other point gives $G(\infty-) = 0$. Hence $G_n(\infty) - G_n(\infty-) \overset{a.s}{\to} 1$, representing a jump. Remembering that for every point $t \in [0, \infty)$ we have $G(t) = 0$ which gives $G_n \overset{a.s}{\to} G = 0$. Keep in mind that G represents the "height" of the curve and so we now have that no jumps occur while $t \in [0, \infty)$, as

well as continuity of L for these values of t.

The empirical probability distribution function is discrete meaning that L_n is piecewise linear and setting $t_n = F_n^{-1}(p)$, the point $(p, L_n(p))$ lies on the segment joining $(F_n(t_n), G_n(t_n-))$ with $(F_n(t_n+), G_n(t_n))$. Consider J, the set of points $p \in (0,1)$ where $F^{-1}(p)$ and $F^{-1}(p-)$ are not atoms of the probability distribution. Then if $p \in J$ there exists an $x_0' < x_0 = F^{-1}(p-)$ and $x_1' > x_1 = F^{-1}(p)$ such that $G(x_0') > G(x_0) - 1/2\epsilon$ and $G(x_1') < G(x_1) + 1/2\epsilon$ for some $\epsilon > 0$. Since x_0 and x_1 are not atoms of the probability distribution then they are the points where $G_n(x_i') \to G(x_i'), i = 1, 2$ and that $G(x_0) = G(x_1) = 0$. Of course then $G_n(x_i') \to 0$. It can be found that there exists a value N such that for $x \in (x_0', x_1')$:

$$L(p) - \epsilon < G_n(x_0') \leq G_n(x-) \leq G_n(x) \leq G_n(x_1') < L(p) + \epsilon, n \geq N$$

We now have that $G_n(x-) \to 0$ and $G_n(x) \to 0$.

By the application of the Glivenko-Cantelli theorem for the case of F^{-1} as in Lemma four then if F^{-1} is continuous at p then $F_n^{-1}(p) \to F^{-1}(p)$ and by Lemma five we get that the accumulation points of the sequence $\{F_n^{-1}(p)\}$ are $F^{-1}(p-)$ and $F^{-1}(p)$.

Therefore we can choose an $n_1 \geq N$ such that $x_0' < t_n < x_1'$ when $n \geq n_1$. Then:

$$L(p) - \epsilon < G_n(t_n-) \leq L_n(p) \leq G_n(t_n) < L(p) + \epsilon, n \geq n_1$$

We have that for every t_n defined by $p \in J$, $G_n(t_n-) \to 0$ and $G_n(t_n) \to 0$ giving $L_n(p) \to 0 = L(p)$. Convergence also holds trivially for $p = 0$ and $p = 1$. We therefore have pointwise convergence. □

9.2 Convergence under J_1 topology

The sequence of curves L_n lie in the space of continuous functions $\mathbb{C}[0,1]$ whereas L lies in $\mathbb{D}[0,1]$, the Skorokhod space which is also the space of all càdlàg functions on $[0,1]$ whereby they are right-continuous with left limits. Therefore, convergence under the uniform metric may not hold i.e $\rho(L_n, L) \not\to 0$). It can be seen that our version of L with $\mu = \infty$ is a càdlàg function due to the jump that takes place at $p = 1$. We can possibly validate convergence in this space under various topologies such as J_1 and M_1. The requirements for the J_1 topology are weaker than uniform convergence in the sense that the jumps of càdlàg paths are accounted for.

Essentially we want $d_{J_1}(L_n, L) \overset{n\to\infty}{\to} 0$ where

$$d_{J_1}(x_1, x_2) := \inf_{\lambda \in \Lambda} \{\|x_1 \circ \lambda - x_2\|_\infty \vee \|\lambda - id\|_\infty\} \tag{19}$$

where the identity map is denoted by $id : [0,1] \to [0,1], id(p) = p$ and
$\Lambda = \{\lambda : [0,1] \to [0,1] : \lambda \in \mathbb{C}[0,1], \text{strictly increasing}\}$.

Goldie proves the convergence of the modified Lorenz curve $\tilde{L}_n \overset{J_1}{\to} \tilde{L}$ when $\tilde{L}(p) = G(F^{-1}(p)), 0 \leq p < 1, \tilde{L}(1) = \tilde{L}(1-) = 1$ where he chooses $\lambda_n = F_n(F^{-1}(p))$ for any discontinuity point p and $\lambda_n = id$ for every continuous point p. This proof can be applied in our case due to the fact that $L(p) = G(F^{-1}(p))$ for all p when concerned with infinite mean giving:

$$\lambda_n(p) = \begin{cases} p, & p \in [0,1) \\ F_n(F^{-1}(p)) = 1, & p = 1 \end{cases}$$

This formulation of λ_n provides immediate uniform convergence to the identity function i.e $\|\lambda - id\|_\infty$ and so now we need only consider $\|L_n(\lambda_n(p)) - L(p)\|_\infty$. The following Lemma asserts that the chosen sequence above is the only possible option when uniform convergence is fulfilled.

Lemma 7. *Suppose $L_n(\lambda_n) \overset{n\to\infty}{\to} L$ uniformly and let p be any point whereby L is discontinuous. Then $\lambda_n(p) = F_n(F^{-1}(p))$*

Proof. This Lemma concerns the points p where G is discontinuous at $t = F^{-1}(p)$ i.e where t is an atom of the distribution function. In our case this would be when $p = 1$. The uniform convergence of G_n to G in the previous Lemma as well as monotonicity gives:

$$\text{if}\, G_n(t_n) \overset{n\to\infty}{\to} G(t) \text{then}\, t_n \geq t \,\text{for large}\, n \tag{20}$$

$$\text{if}\, G_n(t_n-) \overset{n\to\infty}{\to} G(t-) \text{then}\, t_n \leq t \,\text{for large}\, n \tag{21}$$

$L_n(\lambda_n(p)) \overset{n\to\infty}{\to} L(p)$ also means that $G_n(t_n) \overset{n\to\infty}{\to} G(t)$ when $t_n = F_n^{-1}(\lambda_n(p))$. Therefore from (20):

$$F^{-1}(\lambda_n(p)) \geq t \tag{22}$$

and uniform convergence of $L_n(\lambda_n(p))$ to $L(p)$ gives:

$$L_n(\lambda_n(p-)) \overset{n\to\infty}{\to} L(p-) = G(t-). \tag{23}$$

The left side of the previous equation is equal to $L_n(\lambda_n(p)-)$ due to the fact that λ_n is continuous and increasing. If L_n is continuous at λ_n then we would have that $L_n(\lambda_n(p)-) \overset{n\to\infty}{\to} L(p)$ and not $L(p-)$. Hence L_n is discontinuous at $\lambda_n(p)$ as $n \to \infty$. Using the fact that $L(p-) = G(t-)$ where $t = F^{-1}(p)$ on (21) then the left side of the equation becomes $G_n(t_n-)$ and by (22), $t_n \leq t$ for large n. If combined with (23) then $F_n^{-1}(\lambda_n(p)) = t = F^{-1}(p)$. λ_n is a discontinuous point of F_n^{-1} and so $\lambda_n(p) = F_n(F^{-1}(p))$. □

Lemma 8. $d_{J_1}(L_n, L) \overset{n\to\infty}{\to} 0$

Proof. Consider the discontinuous point of L which in our case is when $p = 1$. Then as we already know, $t = F^{-1}(p)$ is an atom of the distribution function and so $\lambda_n(p) = F_n(t)$ and $F_n^{-1}(\lambda_n(p)) = t$ as before. This gives $L_n(\lambda_n(p)) = G_n(t) \overset{n\to\infty}{\to} G(t) = L(p)$. Remembering that:

$$L_n(\lambda_n(p-)) = L_n(\lambda_n(p)-) \tag{24}$$

and using the fact that $L_n(p-) = G_n(t-)$ when $t = F^{-1}(p)$ gives $L_n(\lambda_n(p-)) = G_n(t-) \overset{n\to\infty}{\to} G(t-) = L(p-)$. Using Chung's Lemma on the interval $[0, 1]$ gives the result. □

10 Other inequality measures

In this section we provide a short introduction and formulation of various inequality measurements that are less widely used than the Gini coefficient but nonetheless still important.

10.1 Pietra index

The Pietra index is known under many other names including the Schutz coefficient (Duclos and Araar, 2006) as well as the Hoover or Robin Hood index. It is defined as the proportion of total income or wealth that would have to be redistributed for there to be perfect inequality. Conveniently, we can provide some intuition due to the fact that this is equal to $\max_{0\geq p\leq 1}\{p - L(p)\}$. By taking the derivative of this it is clear to see that this is when the slope of the Lorenz curve is equal to one i.e $L'(p) = 1$. If we have an empirical distribution with n individuals, x_i denoting their wealth or income with $\bar{x}$ being the average then the Pietra index can also be given as:

$$P = \frac{1}{2n}\sum_{i=1}^{n}\frac{|y_i - \bar{y}|}{\bar{y}} \tag{25}$$

10.2 Atkinson index

Atkinson's index is another measure which has the definition (Atkinson, 1970):

$$I_A(\epsilon) = 1 - \left(\int_0^\infty \left(\frac{x}{\mu}\right)^{1-\epsilon} dF(x)\right)^{1/(1-\epsilon)}, \epsilon > 0 \tag{26}$$

The parameter ϵ is known as the "inequality aversion" parameter which controls the weighting applied to values in the distribution. $\epsilon = 0$ assumes no aversion to inequality meaning no social utility is gained from complete reallocation while $\epsilon = \infty$ assumes the opposite. The index lies between zero and one and at a specified ϵ value, a distribution with a higher coefficient is more unequal than a different distribution with a lower index

than this. If we instead have an empirical distribution, we can calculate the index as follows:

$$I_A(\epsilon) = \begin{cases} 1 - \frac{1}{\mu}\left(\frac{1}{N}\sum_{i=1}^{N} x_i^{1-\epsilon}\right)^{1/(1-\epsilon)} & \epsilon > 0 \neq 1 \\ 1 - \frac{1}{\mu}\left(\prod_{i=1}^{N} x_i\right)^{1/N} & \epsilon = 1 \end{cases} \tag{27}$$

Where x_i is the individual wealth or income and μ is the mean.

10.3 Generalised Entropy family

The generalised entropy family of indicators is defined as:

$$I_G(c) = \frac{1}{c(c-1)} \int_0^\infty \left(\left(\frac{x}{\mu}\right)^c - 1\right) dF(x), c \neq 0, 1 \tag{28}$$

Which empirically gives:

$$I_G(c) = \frac{1}{c(c-1)} \frac{1}{n} \sum_{i=1}^{n} \left(\left(\frac{x_i}{\bar{x}}\right)^c - 1\right) \tag{29}$$

At $c = 1$ we get:

$$I_G(1) = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{\bar{x}} \ln\left(\frac{x_i}{\bar{x}}\right) \tag{30}$$

which is the Theil index - usually referred to as the difference between the maximum possible entropy and the observed entropy of the distribution. If we take $\epsilon = 1 - c$ then we obtain the Atkinson index from the generalised entropy index.

Chapter 4 of Lubrano's notes provides a variation of these indices in terms of the Lorenz curve. For the generalised entropy index:

$$I_G(c) = \frac{1}{c(c-1)} \int_0^1 \left([L'(p)]^c - 1\right) dp, c \neq 0, 1 \tag{31}$$

Replacing $\epsilon = 1 - c$ again, for the Atkinson index:

$$I_A(\epsilon) = 1 - \left[\int_0^1 (L'(p))^{1-\epsilon} - 1\right]^{1/(1-\epsilon)}, \epsilon > 0 \tag{32}$$

This means that the indices can be calculated directly from the Lorenz curve.

11 Inequality measurement requirements

Before this we have talked about inequality measures - mainly the Gini index. However, it is necessary for indices such as the Gini index to fulfil some basic requirements to be regarded as an inequality index. These are:

- Anonymity - If the distribution is permuted then the index value is not altered i.e if x' is a permutation of x then $I(x) = I(x')$.
- Population principle - size of the population does not matter i.e $I(x) = I(x, x, \ldots, x)$
- Scale invariance/relative income principle - The values of the distribution do not matter just how they relate to each other i.e $I(\lambda x) = I(x)$ for some $\lambda > 0$.
- Pigou-Dalton principle/principle of transfers - If a transfer of wealth x occurs from one wealthy individual y_j to a less wealth individual y_i and $y_j - x > y_i + x$ then the value of the index falls. In other words if a series of regressive transfers take place then inequality falls.

These requirements are luckily very intuitive and it is easy to see that most indices would fulfil these requirements. The Gini coefficient, Atkinson index and Generalised Entropy indices fulfil these four properties but the Pietra index does not fulfil the principle of transfers (Neves Costa and Pérez-Duarte, 2020).

Additionally we can say that an inequality measure I is Lorenz-consistent if it fulfils the previous statements and if y and z are income or wealth distributions such that if $I(y) > I(z)$ then the Lorenz curve of y lies everywhere to the right of z. If the curves intersect then there are both progressive and regressive transfers that would need to take place to move from one of the curves to the other and so the Lorenz criteria and the Pigou-Dalton principle does not apply. The Gini index fulfils the Lorenz criterion, as well as the coefficient of variation.

12 Pigou-Dalton Monte Carlo simulations

12.1 Flat taxation

We can also introduce the notion of taxation when generating Lorenz curves where we assume our distribution of interest is a wealth distribution. This is an example of a Pigou-Dalton transfer which are mean preserving transfers of wealth, fulfilling the principle of transfers. Following Lubrano's notes in Chapter 4, we can choose a level of taxation (eg 0.25) which we can multiply by the reverse ordered vector of wealth or for the respective population. This corresponds to every individual being taxed 25%. The code below demonstrates this:

```
x = sort(x)
transfers = i * sort(x, decreasing = T)
newx = ((1 - i) * x) + transfers
```

Where x is the initial distribution of wealth and $newx$ is the new distribution after taxation takes place. We take our value of β to be 0.95 instead of 0.6 used before when illustrating convergence. The level of inequality will be lower, therefore producing plots where distinctions between tax rates can be seen easier as the curves will be more

dispersed. Figure four shows a plot of the flat taxation rate imposed on the population and the corresponding Gini indices produced by this, resulting in a U-shaped curve. We use a sample size of one million and repeat the process one thousand times for each taxation level. This is because the Ginis produced are random variables and so some sort of reproducibility is required for good results. No taxation of course results in the distribution of incomes or wealth to be unchanged, retaining the highest level of inequality with a Gini index of around 0.76. As the level of taxation increases, we can see that the amount of inequality present then starts to decrease with it reaching its lowest level at a taxation rate of 40% and 60% at around 0.68 meaning that this level of taxation causes the Gini index to fall of 0.08 - a respectable reduction in inequality for such an imbalanced distribution. It is interesting that at 30% we have higher levels of inequality than at 20% which we would expect to have the opposite relation.

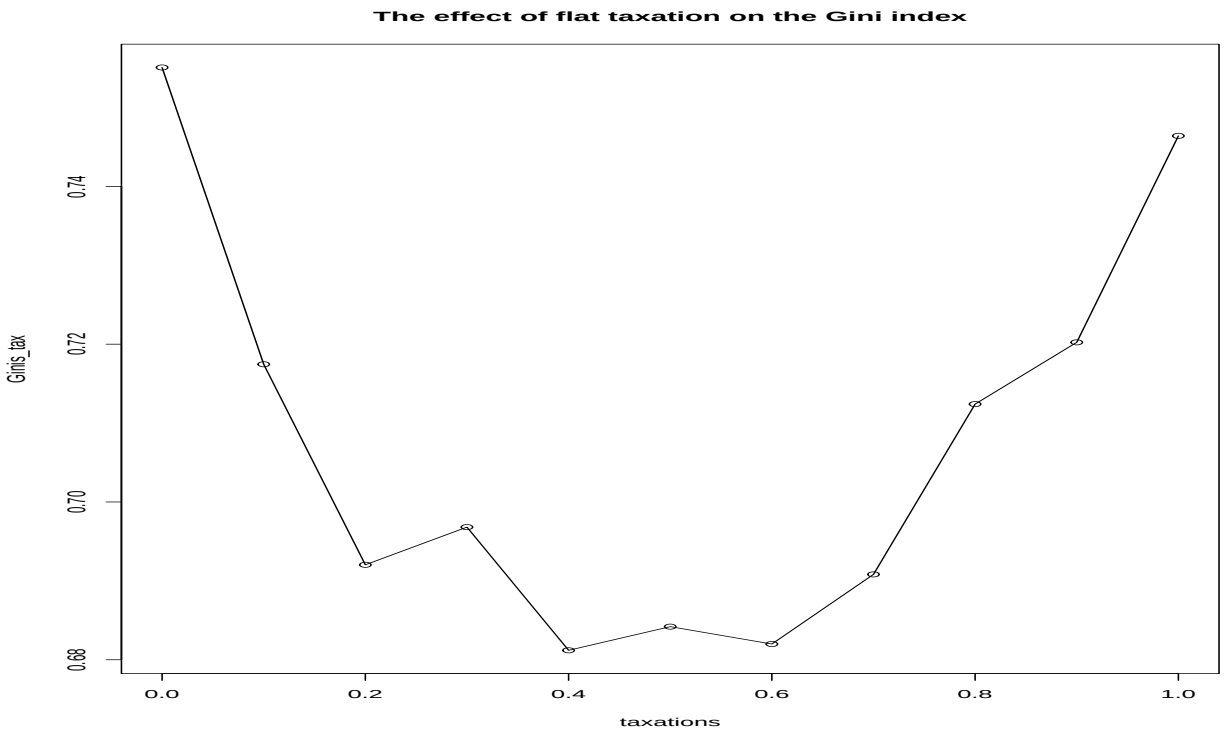


Figure 4: Changes in the Gini coefficient due to a flat taxation level at $\beta = 0.95$

Figure five also shows the Lorenz curves that correspond to these flat rates of taxation, offering another perspective on how inequality is affected by flat taxation.

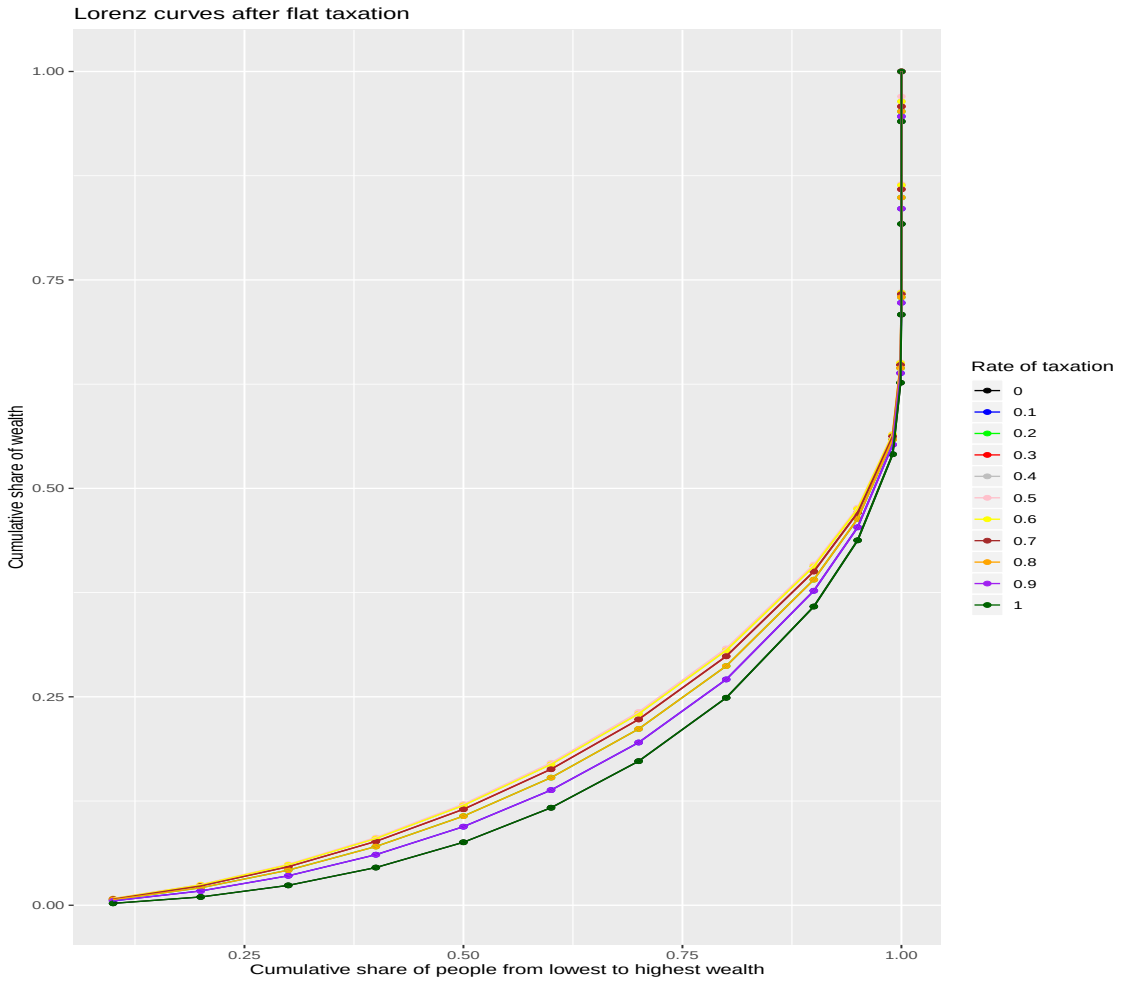


Figure 5: Lorenz curves for each flat taxation rate

12.2 Robin-hood style (quantile) taxation

Using flat taxation is easy to implement and causes a good reduction in the level of inequality but there are more realistic methods we can implement that will hopefully produce a further reduction in inequality. One example of this is a transfer that heavily favours the poor which we will refer to as Robin-hood style taxation. This is where a person is taxed based on which quantile they lie in with the most wealthy individual being taxed 100% and the least wealthy individual not being taxed at all.

Those who lie in the middle of the distribution are still taxed at a 50% rate and so this is why the poor are favoured so largely in this implementation. We can provide a variation of this method by giving some leeway to those being taxed extortionate amounts by taking amounts off of the respective quantiles. For example if we provide a leeway of 0.1, the most wealthy individual will now be taxed less at 90% instead of all of their wealth. We would expect that the first implementation would provide the greatest reduction in inequality out of all of the possible options as this would be the most extreme

version of wealth distribution. This is confirmed in Figure six below whereby we plot the Lorenz curves of each new distribution after redistribution with the Gini indices that correspond to these new distributions displayed in Figure seven. The original distribution pre-tax had a Gini index of around 0.75 meaning that the largest possible reduction in inequality for this example is a fall in the index of around 0.08, offering a slight correction. Hence this method has the same effect as a flat taxation rate of 50%. However, It is surprising that this method did not have a more profound change on the Gini index as there is a large overhaul of the distribution after tax but observing that the distribution is essentially flipped with the rich becoming totally poor and the poor becoming totally rich offers insight into why inequality is not corrected that much.

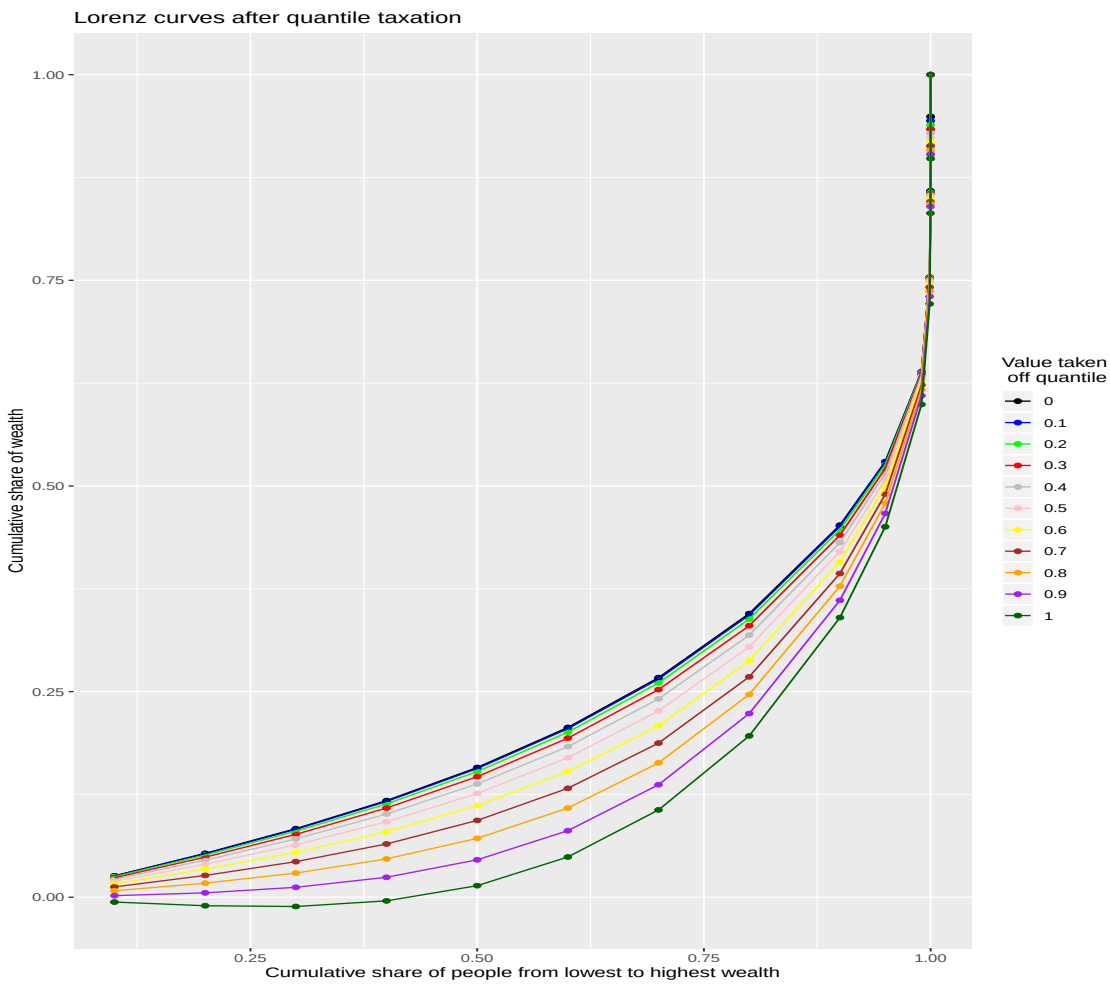


Figure 6: Lorenz curves for quantile taxation at $\beta = 0.95$

We can also produce a plot in the same style as when doing flat taxation whereby we use a sample size of one million and repeat the experiment one thousand times for each level of taxation.

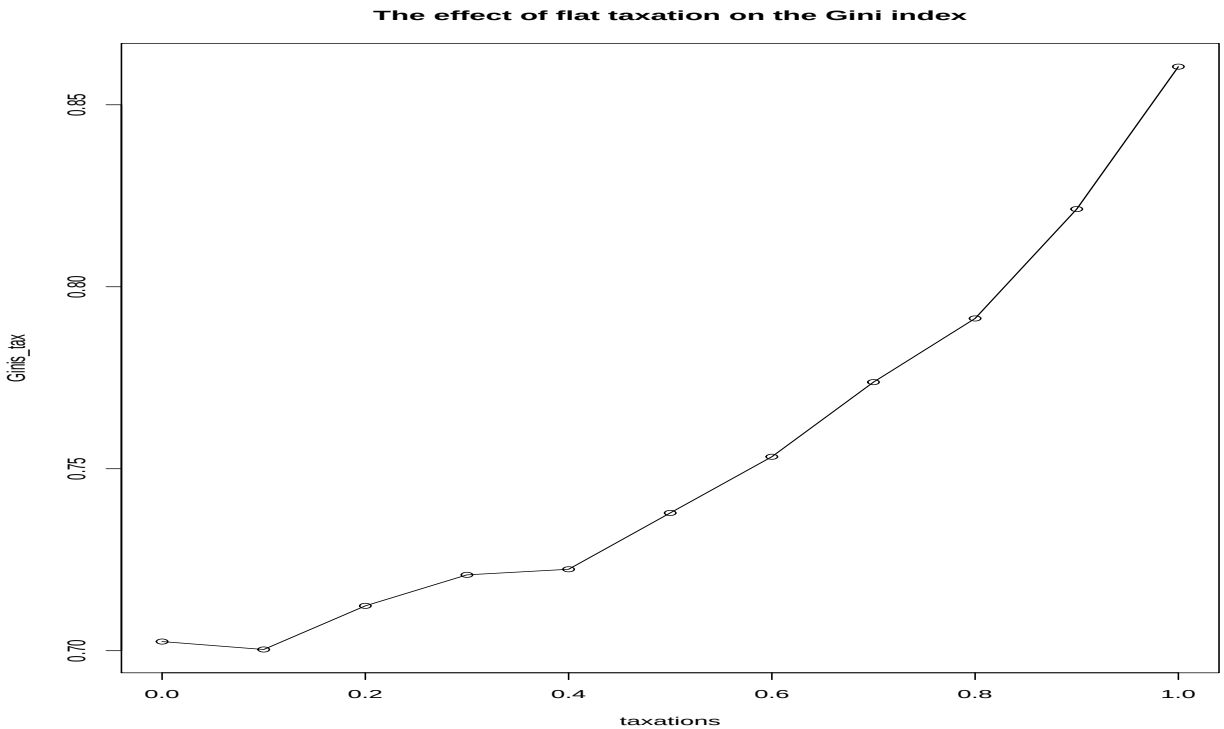


Figure 7: Changes in the Gini coefficient due to variations of quantile taxation at $\beta = 0.95$

The results are as expected where almost no leeway produces the highest reduction in inequality and a leeway of one corresponds to no taxation, therefore retaining the highest Gini.

12.3 Emulating the UK tax system

In the UK the income tax bracket system is organised as described in the table below, with a personal allowance of £12,500 that is not taxable followed by increasingly higher taxation as higher income is achieved. The true mean income in the UK is around £30,000 meaning that the personal allowance is 5/12 of the mean, the basic rate is 5/3 times the mean and so on. Hence we can standardise these brackets for our own distribution and calculate the amount of tax paid by each individual in our distribution, allowing redistribution.

Band	Taxable income	Tax rate
Personal allowance	Up to £12,500	0%
Basic rate	£12,501 to £50,000	20%
Higher rate	£50,001 to £150,000	40%
Additional rate	Over £150,000	45%

Figure eight shows the effect of this taxation style on the Lorenz curve of the distribution where the first type of taxation is none - the original distribution for reference with a Gini index of around 0.75 as before. The second type is the result of redistributing the

taxes back to the poor as we have done in the previous sections whereas the third type is the result of allocating the tax obtained randomly amongst the population. The second type of reallocation has an index of 0.66 - a fall of 0.09. We can conclude that this is a slightly more successful implementation when compared to flat or quantile taxation methods performed previously and produces the largest fall in inequality thus far. The third has an index of 0.74 which is only a reduction of 0.01 - a poor method. However, this method is heavily based on random sampling and so this could be wildly different depending on whether the wealthiest receive their taxed amounts back randomly or not. Instead of using the taxation rates above we could choose different progressive values which may have a larger impact on the reduction of inequality. Keep in mind that the reason that the Lorenz curves for the original distribution and the one produced by the random allocation of taxes intersects is due to the fact that this transfer is not mean preserving - it is not a Pigou-Dalton transfer. The mean income is actually increased which is not reflected in the Gini coefficient but should be considered when choosing the best method for reallocation of wealth.

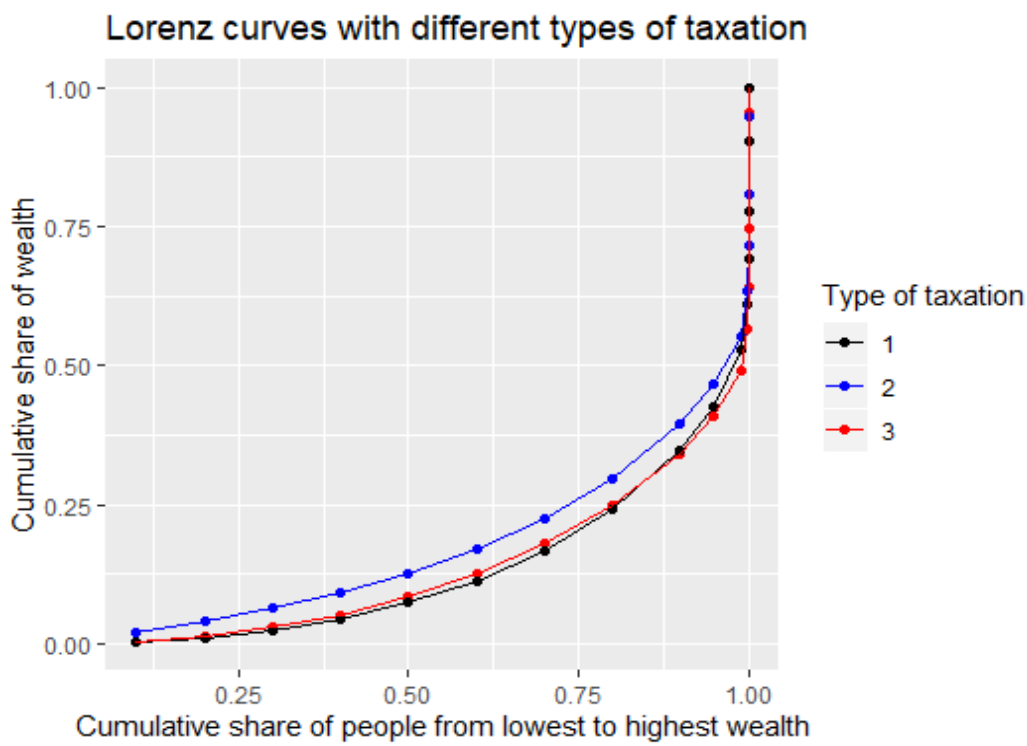


Figure 8: Lorenz curves for emulation of UK tax system at $\beta = 0.95$ where the black line is the original distribution, the red is after taxation and redistribution to the poor while the blue is after taxation and random redistribution.

We then repeated this process 100 times at a sample size of 100000, generating our vector and applying the two different styles of taxation to it. This can be seen in the figure below.

Type of taxation	Gini	Gini sd
None	0.72	0.07
UK system	0.62	0.10
UK system with random reallocation	0.67	0.08

Clearly we can see that the Gini index falls a significant amount of around 0.1 after redistribution according to the UK tax system. Random reallocation has a much less profound effect on the reduction in the Gini index.

Of course in this section we reallocate the taxation amounts directly back to those in the distribution but in reality this is generally not the case, with taxes being used by governments to reinvest back into society while not providing direct relief to those who are considered poor. Hence the reduction in inequality that we see in these simulations will be more profound than in reality. However, this depends on the attitudes of those who receive the tax payments (i.e governments) towards inequality. A more economically liberal government may provide more investment using these tax payments in increasing social mobility through the investment in public sector resources such as education and transport or through more direct methods such as an increase in the minimum wage. Hence it is impossible to say the extent to which taxation will reduce inequality as it largely depends on how these funds are used to address the issue.

13 Generating Mittag-Leffler data and fitting to a power-law distribution

In this section, I aim to generate increasingly large sample sizes of random variables with different parameter values that are Mittag-Leffler distributed. The results can be seen in the eight given plots below, whereby we then go on to fit the tail of this data to that of a power-law distribution. The data plotted below is generated using the previous formula for τ_β.

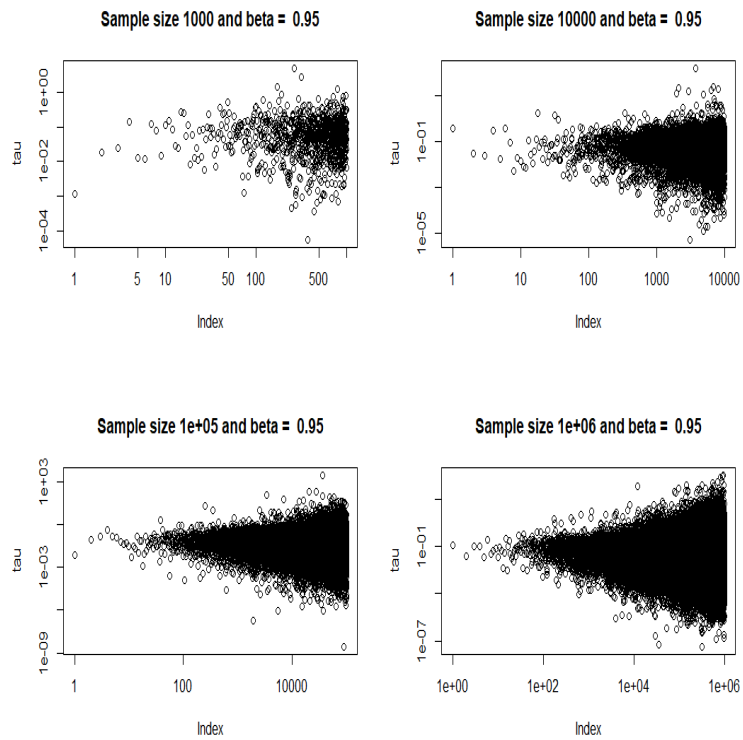


Figure 9: Log-log plot of Mittag-Leffler data with varying sample sizes and tail index $\beta = 0.95$

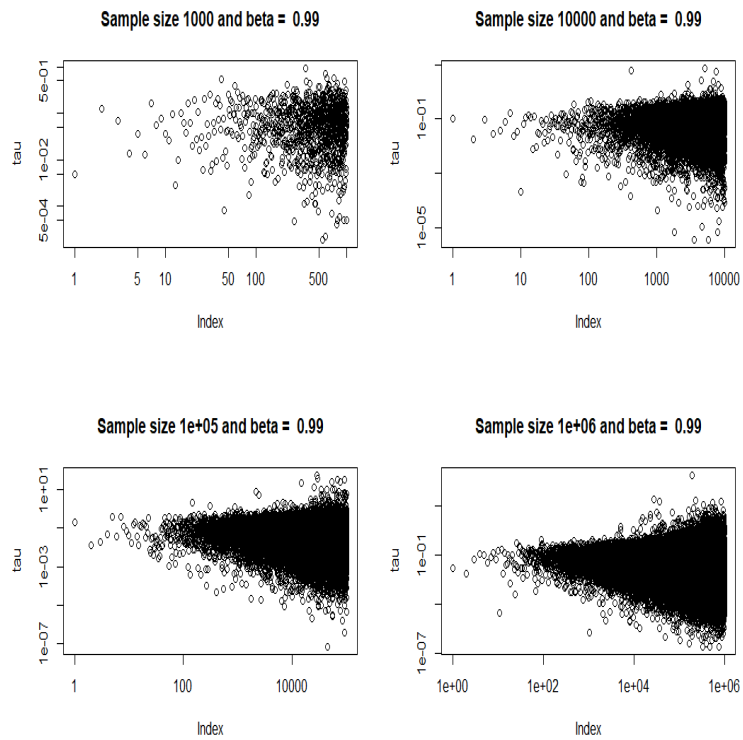


Figure 10: Log-log plot of Mittag-Leffler data with varying sample sizes and tail index $\beta = 0.99$

We now try and fit the tail of this generated data to that of a power-law distribution which has the pdf:

$$p(x) = Cx^{-\alpha} \tag{33}$$

where C is some normalization constant. The complementary cumulative distribution function is also given by:

$$P(x) = \int_x^\infty p(x)dx = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1} \tag{34}$$

The parameters α and $x_{\min}$ need to be estimated. This can be done by finding the value of $x_{\min}$ which minimizes the value of the Kolmogorov–Smirnov test and so is the best parameter when fitting to the data (Lai, 2016).

The process of the implementation is that the unique values in the inputted data are found, which are then looped over and truncated below each value. This is then used in the generation of $x_{\min}$ and then α which is in turn then used for the calculation of the theoretical cumulative distribution function of the power-law data given above. This is then compared to the empirical cumulative distribution function using a KS test as described before.

A vector of the results of the test is then kept, with the smallest value being used to find the corresponding unique value in the data to truncate again, meaning that the parameters that minimise the KS test are chosen.

The data that we pass into the R function is generated from a function that has the

same behaviour as the tail of the Mittag-Leffler function that we previously generated given by equation (31) in (Scalas, Gorenflo and Mainardi, 2004):

$$\Psi(\tau) \sim \frac{\sin(\beta\pi)}{\pi}\frac{\Gamma(\beta)}{\tau^{\beta}}, 0 < \beta < 1, \tau \to \infty$$

using the one parameter Mittag-Leffler function described earlier:

$$E_{\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\beta k + 1)} \tag{35}$$

which depends upon β and has the survival function:

$$\Psi(\tau) = E_{\beta}(-\tau^{\beta}) \tag{36}$$

The value of α generated when fitting the tail index of the data to a power-law distribution uses the previously calculated x_{min} value and is given by the formula which is its maximum likelihood estimation:

$$\hat{\alpha} = 1 + n\left[\sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{\min}}\right)\right]^{-1} \tag{37}$$

We can generate power-law data using the $x_{\min}$ and α parameters we found which can be passed into another KS test (Wash, 2008). The null hypothesis being that the inputted data comes from a power-law distribution and the alternative hypothesis being that it does not, a p-value can be generated. If the p-value is below the criterion of 0.10, we reject the null and produce a value of 1. Up to an index of 2500, these values are summed so that a result of 2500 would mean that at every point the null hypothesis is retained. This hypothesis test allows us to see the extent to which our data is a good fit for a power-law distribution. It is also clear that the relation between α and β is given by:

$$\beta = \alpha - 1 \tag{38}$$

And so it is also true that $\hat{\beta} = \hat{\alpha} - 1$. Below is a table of results showing the produced $\hat{\alpha}$ values corresponding to the specified β values (the tail index).

Sample size	β	$\hat{\alpha}$
1000	0.99	2.014
1000	0.95	2.062
10000	0.99	2.010
10000	0.95	2.051

In every test performed, the null hypothesis is retained and hence our data can be seen as a good fit for a power-law distribution. However, from the table above we can see that we would attain $\hat{\beta}$ values 1.014 and 1.010 for $\beta = 0.99$ and values of 1.062 and 1.051 for $\beta = 0.95$ meaning that we receive a slightly overestimate of β during the fitting with an

error of around 0.1. Additionally, if the fit is accepted we are of course assuming that the mean of the distribution is finite whereas theoretically the data comes from an infinite mean distribution and so this is also a problem. We provide a visual aide below in the form of a log-log plot regarding the fit of the data produced from the survival function acting as the tail of the Mittag-Leffler distribution against a power-law distribution.

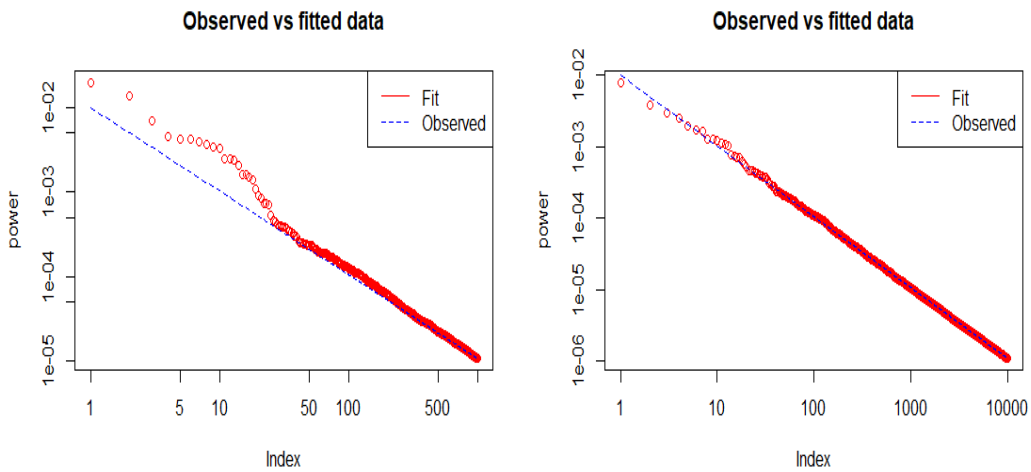


Figure 11: Log-log plot of fit of $\Psi(\tau)$ against power-law data with index = 0.99 and sample size 1000 and 10000

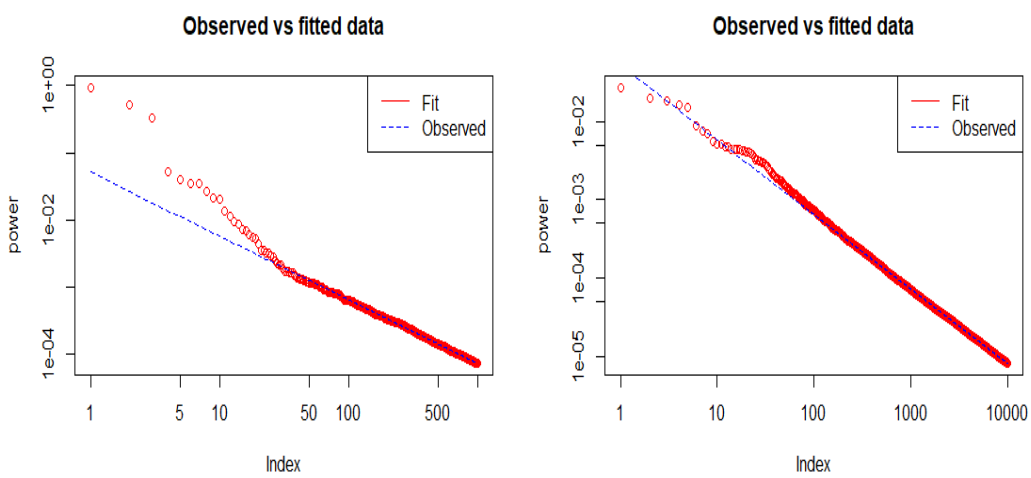


Figure 12: Log-log plot of fit of $\Psi(\tau)$ against power-law data with index = 0.95 and sample size 1000 and 10000

14 Empirical work with the British Household Panel Survey

In this section we explore the first set of data we have which is from the British Household Panel Survey (BHPS) which began in 1991 and contains 18 waves of data. It was obtained from the Institute for Social and Economic Research at the University of Essex. The data includes a large amount of variables regarding general questions about the

concerned household. We make use of a specific variable which is household net income.

Below we offer a plot of the Lorenz curves of household incomes. Despite having the potential to make 18 curves, they are almost indistinguishable and so we only show the last wave. The Gini coefficient that corresponds to this distribution is 0.36.

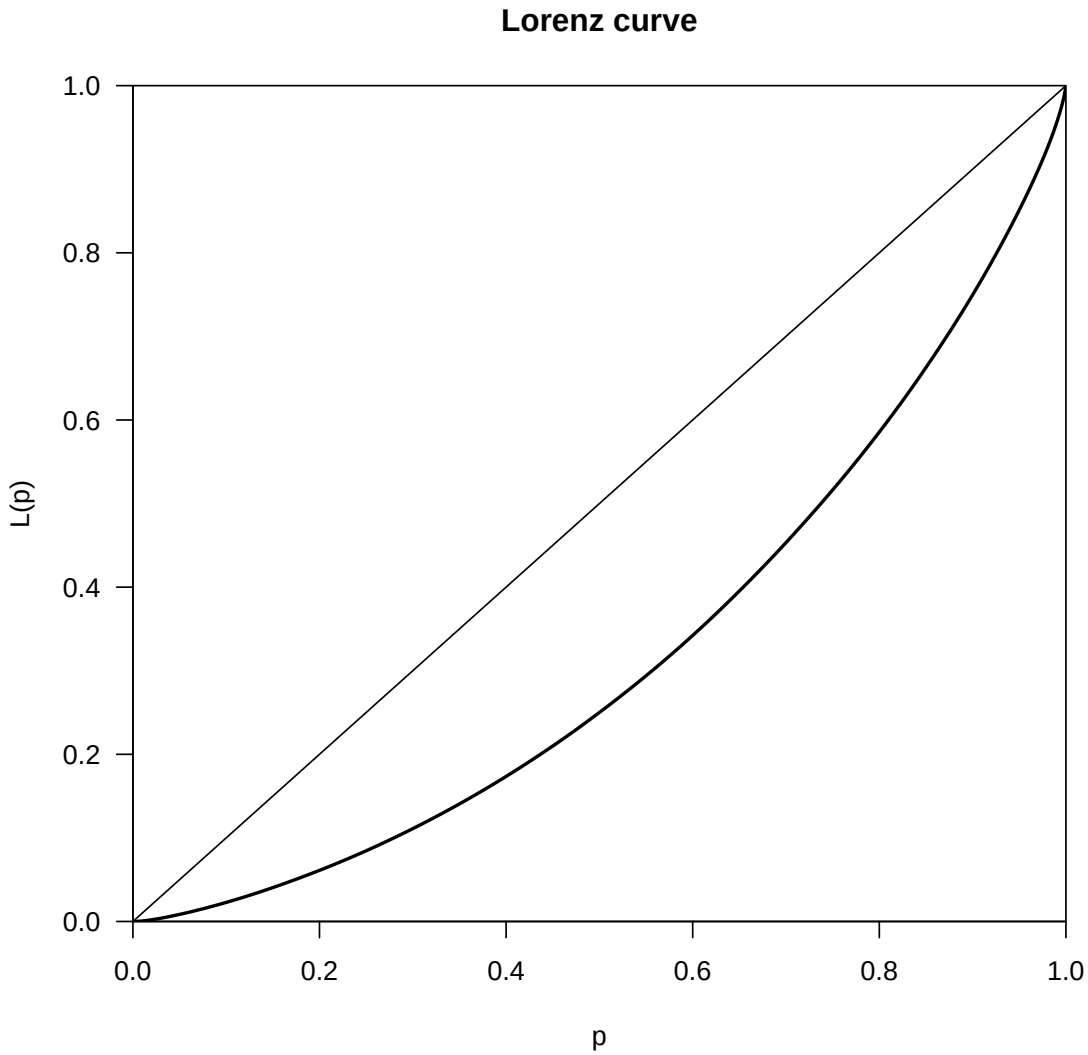


Figure 13: Lorenz curve of household net income for Wave 18

14.1 Fitting to a log-normal distribution

The inequality R package also allows us to plot Lorenz curves easily from various distributions such as the log-normal, Dagum, Singh-Maddala and the Pareto. Below is a fit to the log-normal distribution where the red line corresponds to the parameter 0.45, the blue to 0.55 and the green to 0.6. The green line seems to have the closest resemblance to the Lorenz curve of our data with the log-normal distribution seeming like a good fit overall. Only one parameter is specified which is the log variance due to the fact that the log mean is not relevant to the curve produced due to the scale invariance of the Lorenz

curve. Generally, it seems as though the log-normal is a good potential fit for our data which we will explore further. We also include a table of the Gini indices produced for each wave which will be an important reference when comparing the suitability of various fits and implementations.

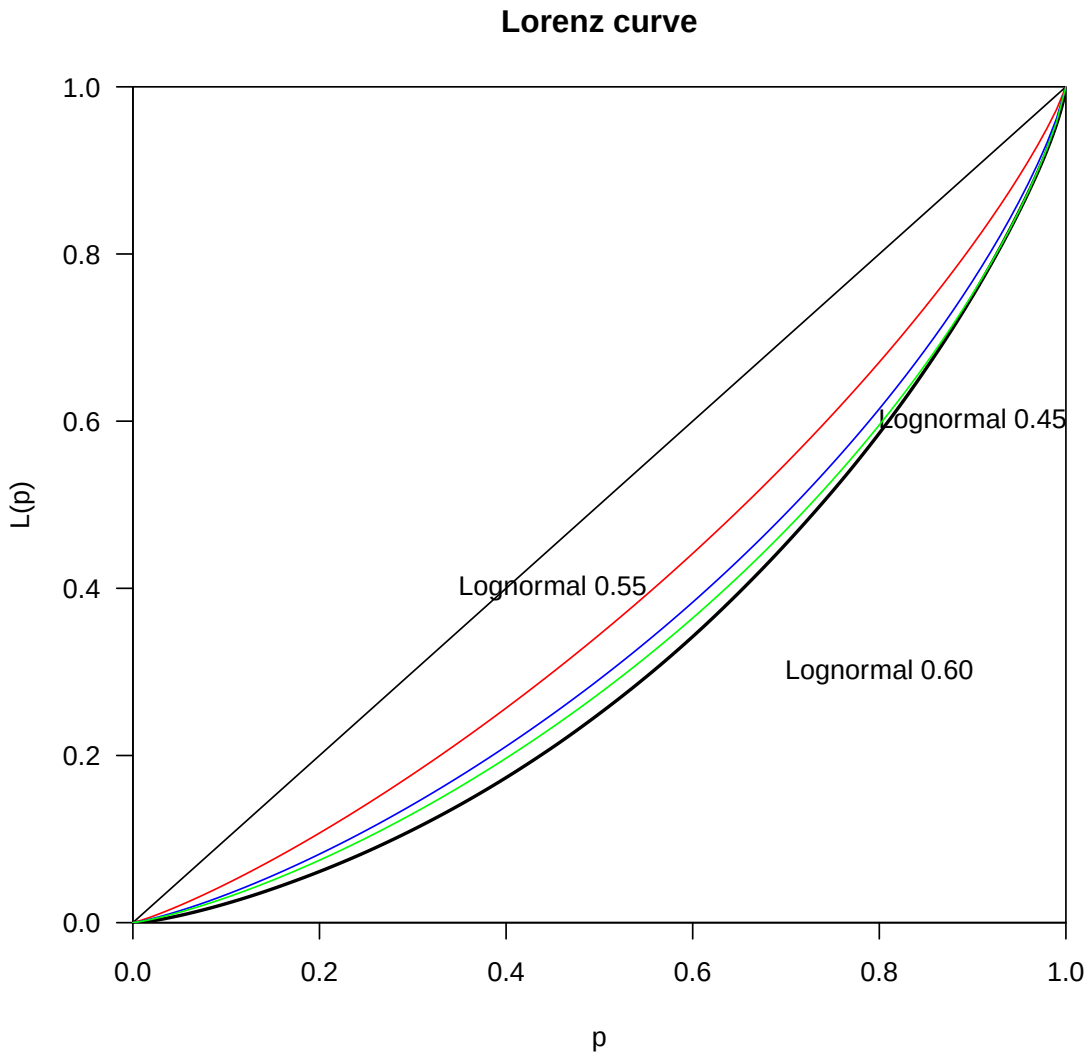


Figure 14: Lorenz curve of household net income wave 18 with fits to a log-normal distribution of various parameters

We now include a table of the true Gini values for each wave. We can produce Gini values of our fit for each wave and then compare them to this table to asses the performance of the fit. Interestingly, it seems as though the trend of inequality is downward - the later waves have less inequality present than in the first waves. This is an odd result to say the least, as we would expect the opposite. However, if we note that this survey only captures low-medium incomes then we can conclude that it is the change in higher incomes that is potentially causing the overall increase in inequality over the years in the UK.

Wave	Gini
18	0.35666
17	0.35554
16	0.35467
15	0.35692
14	0.35981
13	0.36410
12	0.36234
11	0.36266
10	0.36158
9	0.37721
8	0.38503
7	0.37671
6	0.36969
5	0.36945
4	0.36990
3	0.36926
2	0.37450
1	0.37697

Table 1: True Gini values for each wave

The validity of fitting to a log-normal is reinforced by the following figure which shows the density plot of household incomes and a log-normal density in red.

14.2 A Bayesian implementation for the log-normal distribution

We can use Chapters 5 and 8 of Lubrano's notes which provide Bayesian inference on both the log-normal and Pareto distributions in order to find the Gini index as well as its standard error. This can be compared with other methods such as bootstrapping. In particular he states that the posterior density of σ^2 under a non-informative prior is an inverted gamma density. This has n degrees of freedom and a scale parameter $s^2 = \sum(\log(x_i) - \bar{x})^2$. Note that $\bar{x} = \frac{1}{n}\sum \log(x_i)$ in this case. σ^2 values can then be simulated which allows for a direct computation of the Gini index which is $2\phi(\sigma/\sqrt{2}) - 1$ in the log-normal case. Lubrano uses Family Expenditure Survey (FES) data from the UK with waves from 1979-1996 meaning that his implementation will most likely be suitable in our case due to the fact that we also have UK-based data. Also provided are the formulas for various additional indices corresponding to the log-normal distribution which are given in the table below, meaning we can produce standard errors for these as well as the Gini.

Table 5: Various coefficients for the Log-Normal distribution

Coefficient of variation	$\sqrt{\exp(\sigma^2) - 1}$
Lorenz curve	$\Phi(\Phi^{-1}(p) - \sigma)$
Pietra index	$2\Phi(\sigma^2/2) - 1$
Gini index	$2\Phi(\sigma/\sqrt{2}) - 1$
Atkinson	$1 - \exp(-1/2\varepsilon\sigma^2)$
Generalized entropy	$\frac{\exp((\theta^2 - \theta)\sigma^2/2) - 1}{\theta^2 - \theta}$

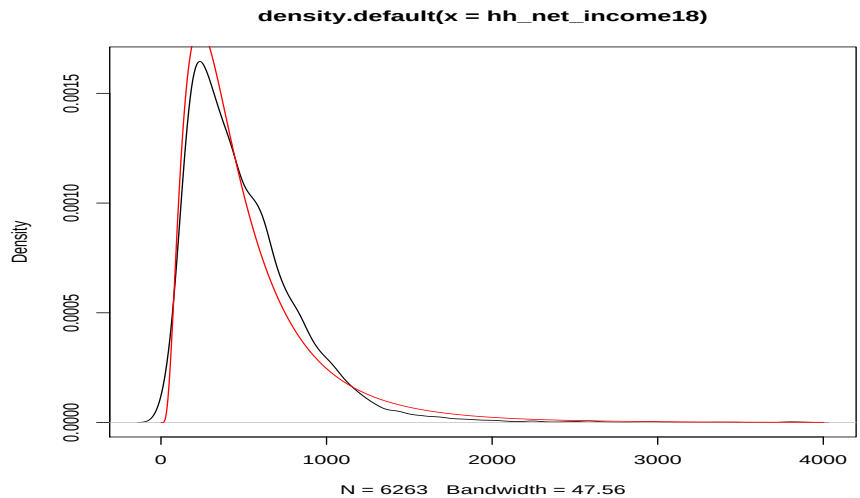


Figure 15: Density plot of Wave 18 of household incomes with a log-normal density plot.

Wave	Gini	Gini SD	Generalised Entropy (0.5)	GE 0.5 SD	Atkinson (0.5)	Atkinson SD	Theil	Theil SD
18	0.39514	0.00327	0.25898	0.00453	0.84548	0.00522	0.24257	0.00396
17	0.39753	0.00319	0.26232	0.00446	0.84165	0.00514	0.24548	0.00390
16	0.38834	0.00310	0.24966	0.00421	0.85623	0.00485	0.23440	0.00370
15	0.39182	0.00304	0.25441	0.00418	0.85075	0.00481	0.23857	0.00366
14	0.39554	0.00302	0.25954	0.00420	0.84484	0.00484	0.24306	0.00367
13	0.39962	0.00301	0.26524	0.00423	0.83828	0.00486	0.24804	0.00369
12	0.40140	0.00300	0.26775	0.00425	0.83540	0.00488	0.25022	0.00370
11	0.40041	0.00278	0.26635	0.00392	0.83701	0.00450	0.24900	0.00341
10	0.41086	0.00308	0.28132	0.00449	0.81987	0.00511	0.26200	0.00388
9	0.41371	0.00306	0.28550	0.00449	0.81512	0.00510	0.26561	0.00387
8	0.41137	0.00363	0.28207	0.00529	0.81902	0.00602	0.26264	0.00457
7	0.40700	0.00357	0.27575	0.00513	0.82623	0.00586	0.25717	0.00445
6	0.40314	0.00389	0.27023	0.00554	0.83256	0.00635	0.25238	0.00481
5	0.41595	0.00409	0.28881	0.00605	0.81139	0.00684	0.26845	0.00521
4	0.41040	0.00395	0.28067	0.00574	0.82062	0.00654	0.26143	0.00496
3	0.40376	0.00394	0.27112	0.00561	0.83154	0.00643	0.25315	0.00487
2	0.41285	0.00399	0.28426	0.00584	0.81655	0.00663	0.26453	0.00504
1	0.41360	0.00384	0.28534	0.00564	0.81531	0.00640	0.26546	0.00487

Table 2: Various inequality indices and their standard errors for all of the data when fitting to a lognormal distribution.

The changes in Gini values over the waves is consistent with the true change in Gini values but with an overestimate of around 0.04 for each, hinting towards to log-normal possibly not being the ideal fit. Despite this, the standard errors are all low. All of the

other indices change in accordance with the Gini with a fall over time with an exception from the Atkinson which rose instead.

What may be a better fit than using a simple log-normal distribution is using a mixture of two log-normal distributions using the *mixdist* R package. Results can be seen below, where we specify the means to be 300 and 600 and the standard deviations to be 100 and 200. This produces a marginally better fit than before seen below:

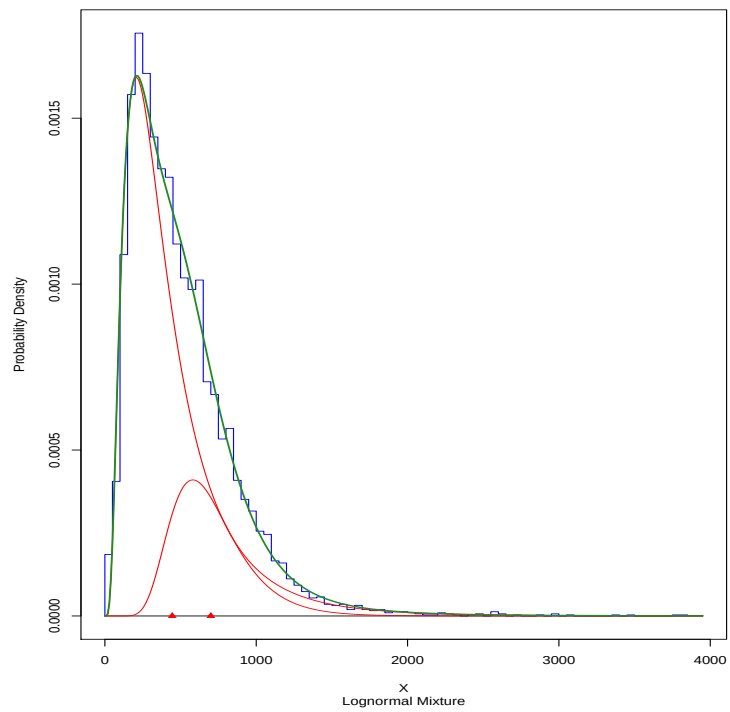


Figure 16: Density plot of Wave 18 of household incomes with a mixture of two log-normal densities.

14.3 Bootstrapping

As an alternative to the Bayesian implementation, we can use a bootstrapping method instead where we can once again find the mean Gini index and the standard error. This is given in the table below.

Wave	Mean Gini	Standard deviation
18	0.35659	0.00324
17	0.35548	0.00325
16	0.35462	0.00320
15	0.35683	0.00311
14	0.35977	0.00369
13	0.36406	0.00368
12	0.36226	0.00345
11	0.36267	0.00291
10	0.36150	0.00322
9	0.37715	0.00421
8	0.38481	0.00618
7	0.37664	0.00451
6	0.36958	0.00417
5	0.36933	0.00447
4	0.36987	0.00417
3	0.36917	0.00384
2	0.37434	0.00392
1	0.37690	0.00370

Table 3: Bootstrapped Gini indices for Waves 1-18

Clearly the bootstrap method is a much better method for producing accurate values of the Gini compared to the Bayesian log-normal fit. For every wave, the bootstrapped Gini values are very close to their true counterpart, with almost no discernable difference. We also include a plot of the bootstrapped values as well as a similar plot for the log-normal Bayesian values for wave 18.

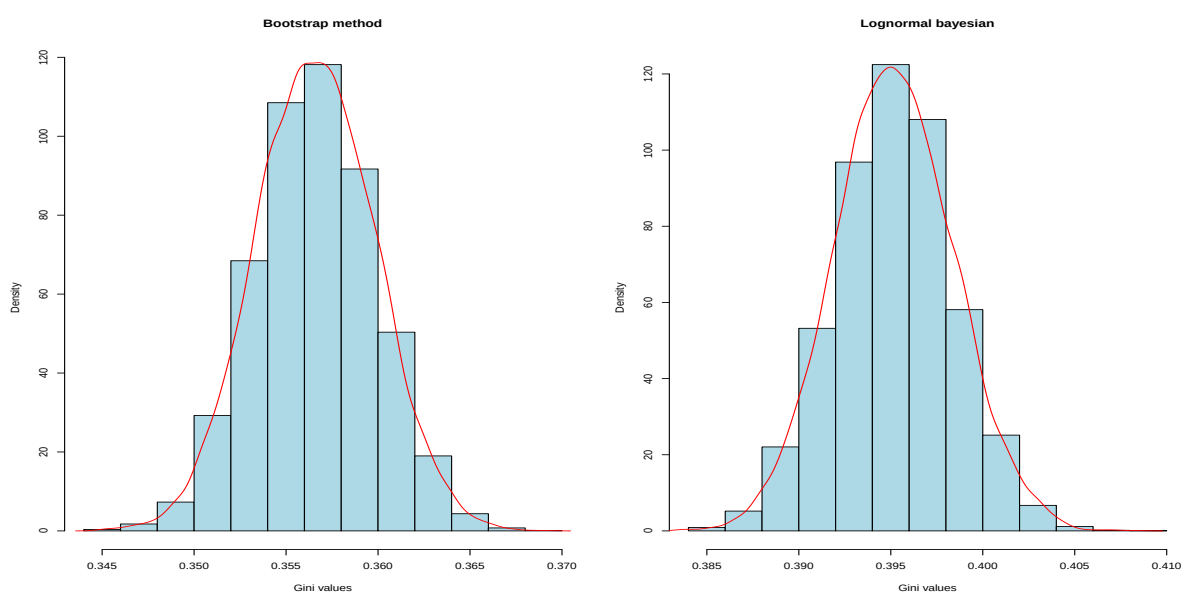


Figure 17: Plot of 10000 bootstrapped and log-normal fitted Gini indices for Wave 18

As the figure above shows, the spread of the two implementations is similar, with low variance. But as claimed before, the bootstrapped values are much closer to that of the true values and so we prefer this method.

14.4 A Bayesian implementation for the Gamma distribution

We now aim to provide an original result whereby we fit to a Gamma distribution using a Bayesian implementation instead. We first find the maximum likelihood estimator of the shape parameter k when the pdf is given in its (k, θ) parametrisation:

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \tag{39}$$

for $x > 0$ and $k, \theta > 0$ and where $\Gamma(k)$ is the gamma function given by $\Gamma(\alpha) = (\alpha - 1)!$

The log-likelihood of the Gamma distribution is given by:

$$(k-1)\sum_{i=1}^{N} \ln(x_i) - \sum_{i=1}^{N} \frac{x_i}{\theta} - Nk\ln(\theta) - N\ln(\Gamma(k)) \tag{40}$$

Taking the derivative and setting to zero gives the maximum likelihood estimator of the θ parameter which is $\hat{\theta} = \frac{1}{kN}\sum_{i=1}^{N} x_i$. This can be substituted back into the log-likelihood giving:

$$(k-1)\sum_{i=1}^{N} \ln(x_i) - Nk - Nk\ln(\frac{\sum x_i}{kN}) - N\ln(\Gamma(k)) \tag{41}$$

Which when taking the derivative and setting to zero again gives:

$$\ln(k) - \psi(k) = \ln\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) - \frac{1}{N}\sum_{i=1}^{N} \ln(x_i) \tag{42}$$

where $\psi(k)$ is the digamma function which is given by $\psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$. By letting:

$$s = \ln\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) - \frac{1}{N}\sum_{i=1}^{N} \ln(x_i) \tag{43}$$

then k can be approximated by:

$$k \approx \frac{3 - s + \sqrt{(s-3)^2 + 24s}}{12s} \tag{44}$$

which is within 1.5% of the correct value (P. Minka, 2002). Hence for each wave of data we can find s, calculate k and then simulate Gamma random variables with the shape parameter which is k which we just found. The reason we use a Gamma distribution again is due to the fact that it is conjugate with itself. The Gini index can then be found which has the form:

$$G = \frac{\Gamma\left(\frac{2k+1}{2}\right)}{k\Gamma(k)\sqrt{(\pi)}} \tag{45}$$

For Wave 18 we tend to get a Gini index of around 0.4, an overestimate of 0.04 when compared to the true Gini value. The log-normal fit produced a slightly closer estimate and so we prefer that method instead. In addition, the standard error produced for this simulation is 0.12 which is much too large for this fit to be accepted and provides further support to the log-normal fit instead. The code used is:

```
BayesGiniGamma = function(x,ndraw){
  # Bayesian SD based on the Gamma fit
  n = length(x)
  x = x[x > 0 ]
  s = log((1/n)*sum(x)) - (1/n)*sum(log(x))
  k = ((3-s) + sqrt((s-3)**2 + 24*s))/(12*s)
  knew = rgamma(ndraw, shape = k, scale = 1)
  G = (gamma((2*knew + 1)/2))/(knew*gamma(knew)*sqrt(pi))
  return(c(mean(G), sd(G))

}
```

15 Empirical work with Office of National Statistics data

In this section we perform some analysis on the second set of data which was taken from the Office of National Statistics' Wealth in Great Britain Wave 4. It includes data from 2006-2014 with two year intervals each containing two columns. The first column is cumulative households and the second is cumulative wealth. This differs from the BHPS data in the sense that wealth is now being studied rather than incomes. Therefore we expect greater levels of inequality due to the fact that the distribution of wealth is intrinsically more unequal than that of incomes. This can be attributed to the fact that wealth can be increased more easily due to the cumulative effect that it has. However, the level of income an individual has clearly contributes to their wealth and so there is a relation.

15.1 Lorenz curves and data manipulation

In subsequent plots we label each period by the year that the period started. For example, if we are concerned with 2006-2008, this will be referred to as the 2006 data. The cumulative nature of the data allows us to directly plot the Lorenz curve for each period which is given below:

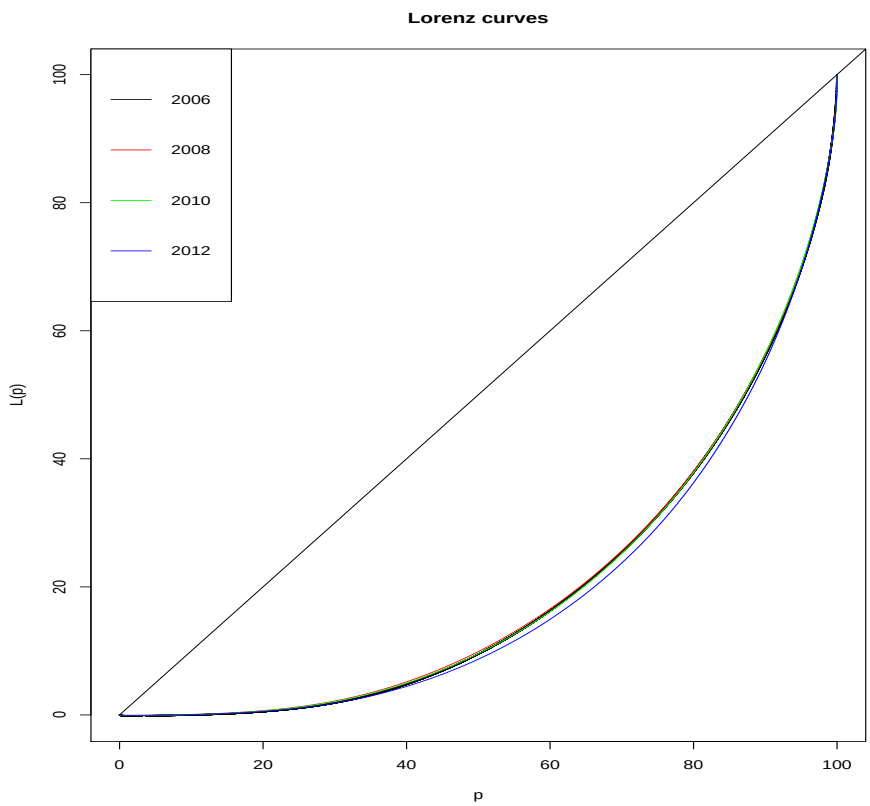


Figure 18: Lorenz curves of total wealth in the UK from 2006-2014

The Lorenz curves for this data are as expected, moving further to the right as a result of increased inequality every two years. The lines are very bunched and hard to see due to the fact that almost no discernable change in the curves occurred over the years. However considering the relatively small time-frame that the data was taken over, there is a respectable increase in inequality. Notice that the curves seem to be much more further to the right than the curves generated previously using the BHS data. This is most likely because household surveys rarely interview the lavishly rich or even those in the top-decile, offering a skewed middle class representation of wealth. Hence it is safe to assume that the BHS data mostly consists of middle and lower incomes. However our new data on total wealth is more comprehensive and takes into account the full scope of households - it is more representative of the true inequality present in the UK.

Despite being able to plot the Lorenz curve directly, we can't do much more with the data in its current format. Another document provided by the Office of National Statistics provides the total wealth for each of the periods and so if we a total number of households we can manipulate this data to produce the level of wealth each household holds - a distribution similar to the BHPS data.

Due to the scale invariance of inequality measures, the number of households and wealth introduced do not affect these. The transformations applied use the diff() function in R

to reverse the cumulative nature of the data to be in a format showing what percentage of households own a certain percentage of total wealth. The total wealth introduced then allows a distribution of unique wealths to be produced. The rep() function then allows the replication of a certain value of wealth that a certain number of households own. These vectors can all be concatenated to produce the final distribution.

The previous log-normal Bayesian implementation was tried for this data but the addition of high incomes resulted in a poor fit as the log-normal is best suited towards medium incomes such as the previous set of data. In addition, Lubrano provides a Bayesian method but for the Pareto distribution - a better fit for higher level incomes such as this set of data. Unfortunately though, this still does not provide a good fit with high standard errors. Therefore we work in favour of a bootstrap method as done before. This allows the production of the following table, where the Gini values are the true values corresponding the data. The bootstrap values are almost exactly the same as the true values, most likely due to the large sample size of the vectors produced ($\sim$ 100,000 for each time frame).

Years	Gini index	Standard error	95% confidence interval	Bias
2006-2008	0.583	0.00168	[0.5800, 0.5866]	7.05e-05
2008-2010	0.587	0.00172	[0.5832, 0.5900]	-8.17e-05
2010-2012	0.611	0.00207	[0.6074, 0.6155]	3.66e-05
2012-2014	0.621	0.00193	[0.6176, 0.6252]	6.931e-05

15.2 First order stochastic dominance

We can also produce dominance curves of order one that correspond to this data. Graphically, dominance curves give a proportion of the poor for a given level of z, a chosen income. The probability distribution F stochastically dominates the distribution G at the order one iff:

$$F(x) < G(x) \forall x \in [0, \infty) \tag{46}$$

We can produce the following plot again through the use of the *ineq* package in R, with specific use of the *Foster* function. We can specify the poverty measure, where one corresponds to the headcount ratio and two is the poverty gap ratio (Foster, Greer and Thorbecke, 1984). This family of poverty indices is given by:

$$P_a = \int_0^z (1 - x/z)^a f(x) dx \tag{47}$$

where f is the distribution, z is the poverty line and a is the specified poverty measure. $a = 0$ corresponds to the headcount ratio and first order stochastic dominance, with

$a = 1$ being the poverty gap and concerning second order stochastic dominance which will be explored later. For $a = 0$ we have:

$$P_0 = \int_0^z f(x)dx \tag{48}$$

If we multiply the number of households by P_0 then we can obtain the number of poor people at a given z. For $a = 1$:

$$P_1 = \int_0^z (1 - x/z)f(x)dx \tag{49}$$

This index fulfils the principle of transfers while the headcount ratio does not.

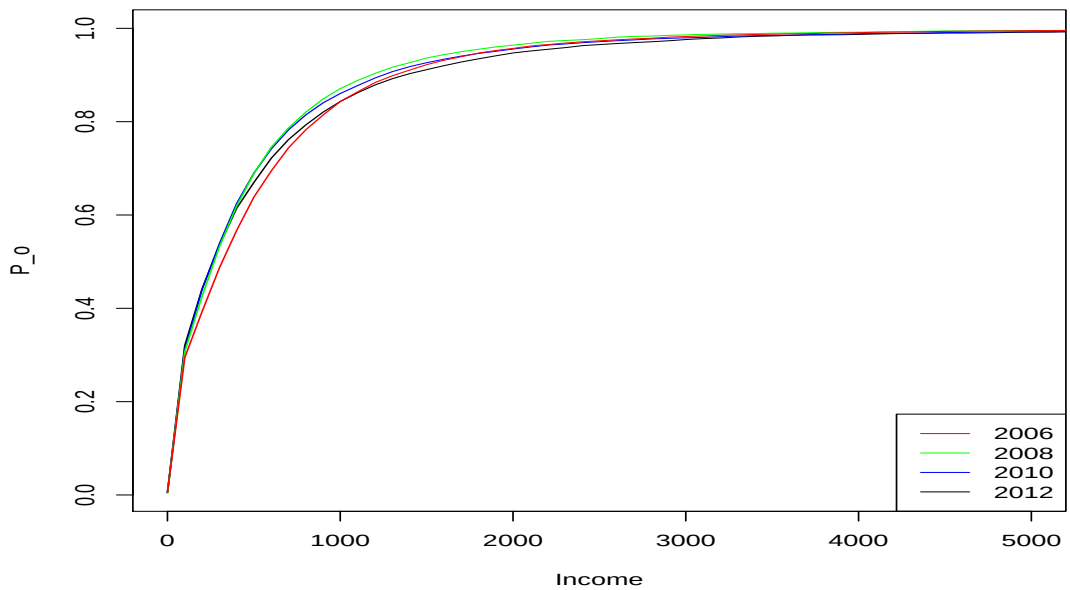


Figure 19: Dominance curves of total wealth in the UK from 2006-2012 restricted to the range £0-£5000 at order one

The elements in the produced distributions corresponding to each two year period are dependent on the number of households to choose. A realistic amount of households ($\sim$ 10,000,000) would generate a vector that is too long and so this had to be scaled down. The number of households chosen is one million so the values of wealth in our distributions are likely to be scaled down representations. This is why the wealth values seem to range mostly in the £0-£5000 range resulting in an overall mean of £510. The plot is restricted from the £0-£5000 range due to the fact that are some huge outliers with large incomes (up to £170,000) and so these are left out of the visualisation. These outliers represent the ultra-wealthy and are the reason for the higher level of inequality when compared to the previous set of data.

Overall from the plot we can see that all of the curves are relatively close to each other, representing a relatively insignificant change in the proportion of the poor population

from 2006-2014. Surprisingly the 2012 distribution dominates the 2008 and 20010 distributions despite an increase in inequality over the total time period. However, the red and black lines representing 2006 and 2012 respectively cross at around the £1000 point and so it is harder to compare these two. We now offer a close up view of the curves for wealths in the range £0-£500 so that we can further investigate the crossing in curves from 2006 and 2012.

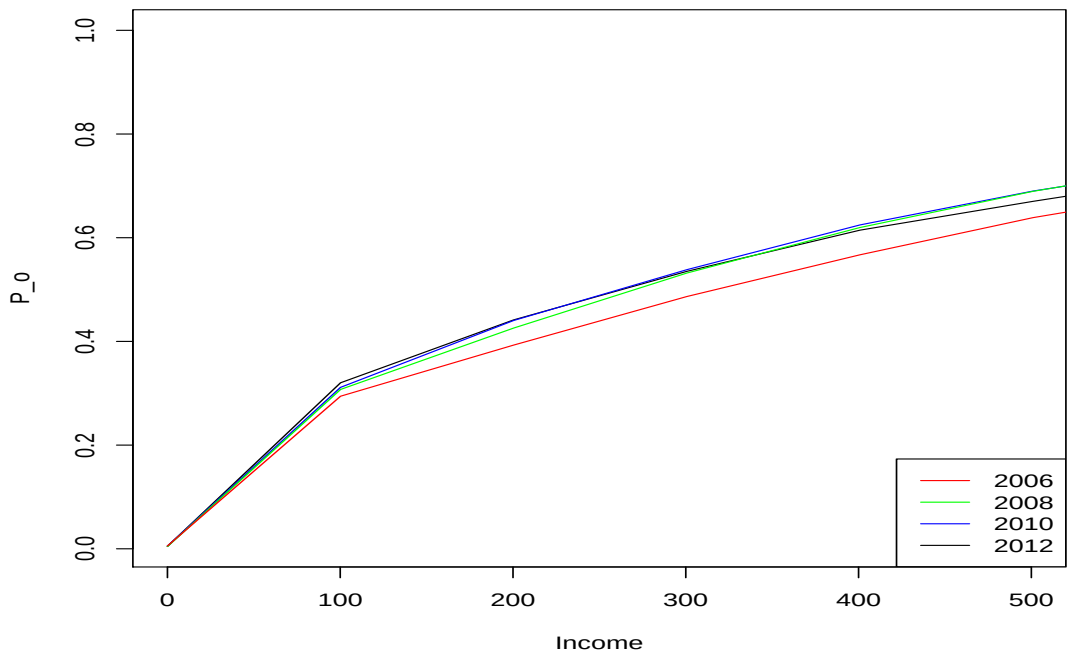


Figure 20: Dominance curves of total wealth in the UK from 2006-2012 restricted from incomes in the range £0-£500 at order one

It seems as though the situation is different at the £0-£500 range. For the very poor, the situation became slightly worse from 2006-2014 with each former year group dominating the next as we would expect in accordance with the rise in inequality during this period. This could mean that potentially the very poor are the group that are most negatively affected by the increase in inequality from 2006-2014.

15.3 Stochastic dominance t-test

We now produce a t-statistic with first order non-dominance as the null hypothesis (Davidson and Duclos, 2013). If we consider the cumulative distribution functions F_A and F_B with respective sample sizes N_A and N_B then the t-statistic is:

$$t = \frac{\sqrt{N_A N_B}(\hat{F}_A(z) - \hat{F}_B(z))}{\sqrt{N_B \hat{F_A}(z)(1 - \hat{F}_A(z)) + N_B \hat{F}_B(z)(1 - \hat{F}_B(z))}} \tag{50}$$

with z being a set of equidistant points which we can generate using seq() in R. Below is a table of the respective means for each distribution along with each poverty line defined as half the mean.

Years	Mean wealth	Poverty line
2006-2008	509.86	254.93
2008-2010	481.88	240.94
2010-2012	544.61	272.30
2012-2014	675.00	337.50

Table 4: Poverty indices

Hence a larger value favours the alternative hypothesis that dominance is present with a large positive statistic indicating that the dominance curve at that particular point for the former date is higher than the latter. The table below illustrates the results of the t-test in the £0-£5000 range.

Wealth	2006-2012	2006-2008	2006-2010	2008-2010	2008-2012	2010-2012
0	-2.97	-0.68	-0.45	0.22	-2.26	-2.48
500	-14.37	-23.39	-23.68	-0.27	8.87	9.13
1002	0.03	-17.22	-10.49	6.66	17.02	10.35
1503	8.49	-12.13	-3.41	8.62	20.31	11.72
2004	9.92	-7.78	1.44	9.11	17.42	8.36
2505	10.20	-6.67	2.53	9.08	16.57	7.57
3006	9.88	-5.90	4.65	10.39	15.47	5.17
3507	6.67	-4.21	4.74	8.81	10.69	1.92
4008	7.82	-2.27	5.04	7.20	9.92	2.76
4509	8.62	-2.29	6.52	8.66	10.70	2.10

Table 5: Stochastic dominance t-test at first order

We have stochastic dominance at almost every value of z and that the table corresponds to the previous plot. It is clear from the table that every other period seems to dominate the 2008 distribution at almost every point. This can possibly attributed to the economic recession that took place during this period whereby a larger proportion of households lived under the relative poverty line. This can be reinforced by the table including the means and poverty lines above by noting that the mean wealth in 2008-2010 was much lower than in the other periods. Particularly, it is 30% less than in 2012-2014, a significant difference. Earlier it was stated that the situation got worse for the poor but if we consider under the poverty line by looking at the first row shows little evidence for dominance. Instead it is around the £500 mark where dominance occurs, suggesting those with medium wealths were the most affected.

15.4 Second order stochastic dominance

The probability distribution F stochastically dominates the probability distribution G at the order two iff:

$$\int_0^x [F(t) - G(t)]\, dt < 0 \forall x \in [0, \infty] \tag{51}$$

We can plot dominance curves of order two which we do below:

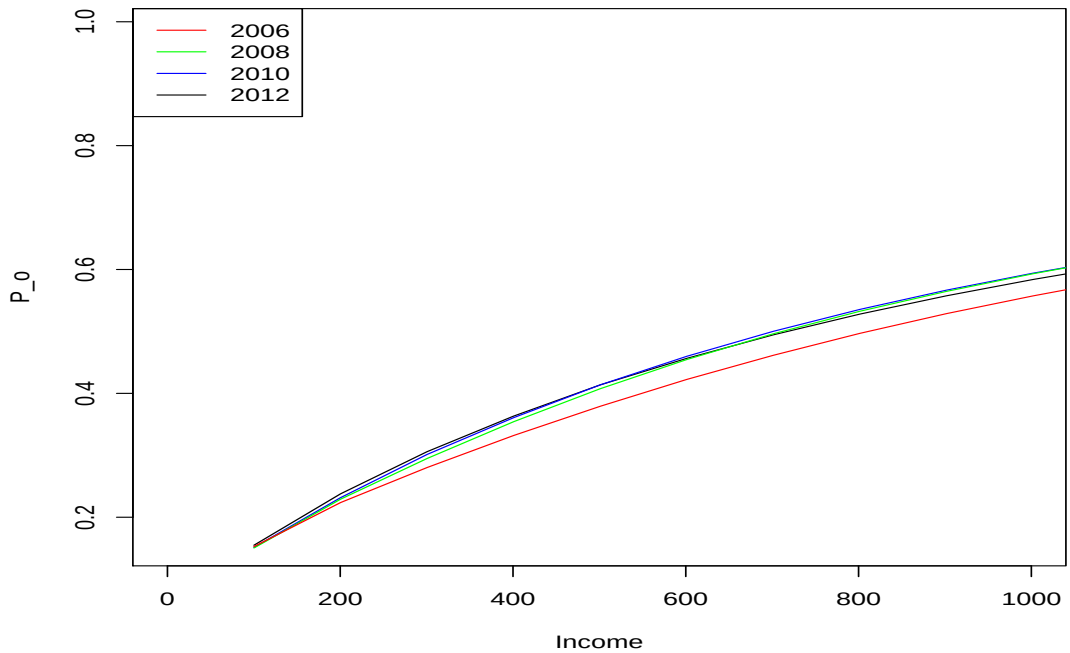


Figure 21: Dominance curves of total wealth in the UK from 2006-2012 restricted from incomes £0-£1000 at order two

The reason for being concerned with second order stochastic dominance is that dominance of first order is not comparable when two curves intersect as we have especially with the 2008, 2012 and 2010 distributions. The surfaces under the cumulative distribution functions are computed, potentially removing the intersection. Clearly the 2006 distribution stochastically dominates all of the other years at the second order in this range, with the dominance curves of the other three distributions being very close with multiple intersections. Hence it does not seem as though second order dominance is as useful as previously thought.

However, second order dominance is relevant due to the fact that it is equivalent to Lorenz domination. We say that F Lorenz dominates G in the Lorenz sense iff:

$$GLC_F(p) \geq GLC_G(p) \forall p \in [0, 1] \tag{52}$$

Essentially this means that the generalised Lorenz curve of F lies everywhere above that of the curve produced by G. We include the proof which is taken from Chapter 7 of Lubrano's notes.

Conjecture 9. *Let us consider two income distributions F and G. It is strictly to say that F stochastically dominates G at the second order or to say that F Lorenz dominates G in the Generalized Lorenz sense.*

Proof. We need to show that:

$$\int_0^z [G(t) - F(t0]dt \geq 0 \text{ for all } z \iff \int_0^p [F^{-1}(q) - G^{-1}(q)]dq \geq 0$$

We have:

$$\int_0^z [G(t) - F(t)]dt = \int_0^z [(1 - F(t)) - (1 - G(t))]dt$$

Using integration by parts with $u = t$ and $v = -(1 - F(t))$ we get:

$$\int_0^z (1 - F(t))dt = \int_0^z tf(t)dt$$

Let $t = F^{-1}(q)$ with $dq = f(t)dt$, then:

$$\int_0^z (1 - F(t))dt = \int_0^z tf(t)dt = \int_0^p F^{-1}(q)dq$$

□

We can verify this relation with our data by plotting the Generalised Lorenz curves and comparing the ordering of the curves with the previous plot regarding second order dominance.

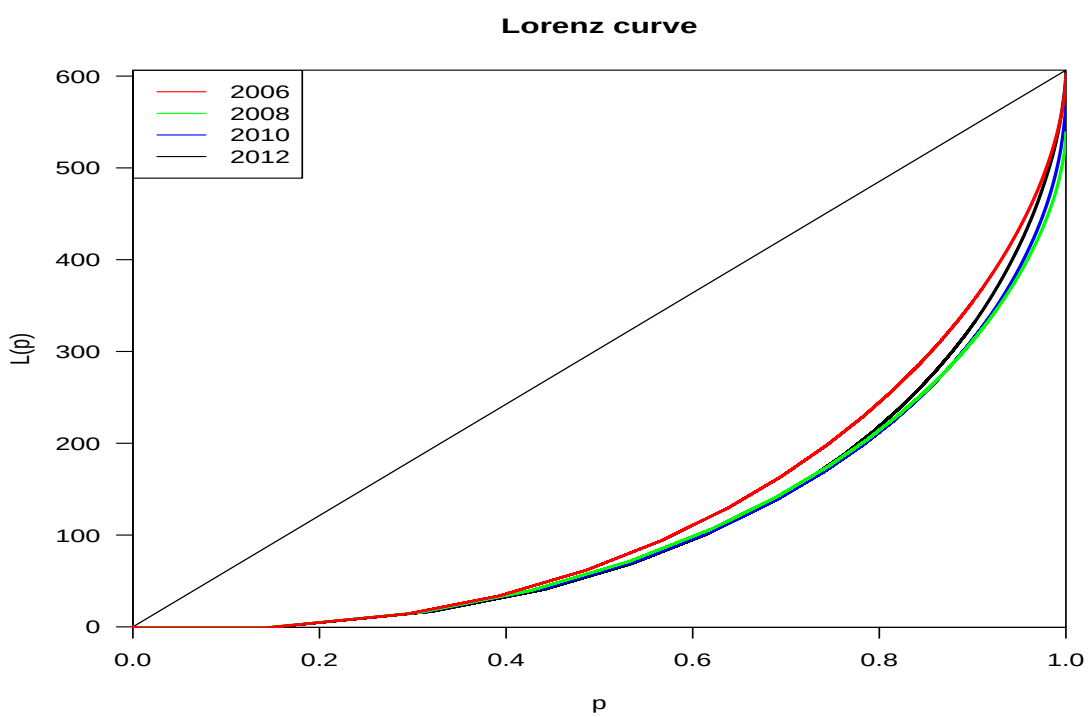


Figure 22: Generalised Lorenz curves of total wealth in the UK from 2006-2012.

It is easy to see that the 2006 distribution is preferred to all of the others in the Generalised Lorenz sense as we would expect with the ordering of the curves remaining the same across the plots. The 2008, 2010 and 2012 distributions again intersect often and

so it is hard to conclude which are preferred in the Generalised Lorenz sense between these three. Hence this provides a visual confirmation of the previous equivalence.

Due to the crossing of dominance curves, it can be hard to tell which have less inequality present. Therefore we now offer a plot of the Atkinson index for varying epsilons. According to Atkinson, empirical studies helped him conclude that the Gini is sensitive to changes in the middle class in accordance with $\epsilon < 1$. $\epsilon > 1$ results in greater concern for the poor (Atkinson 1970).

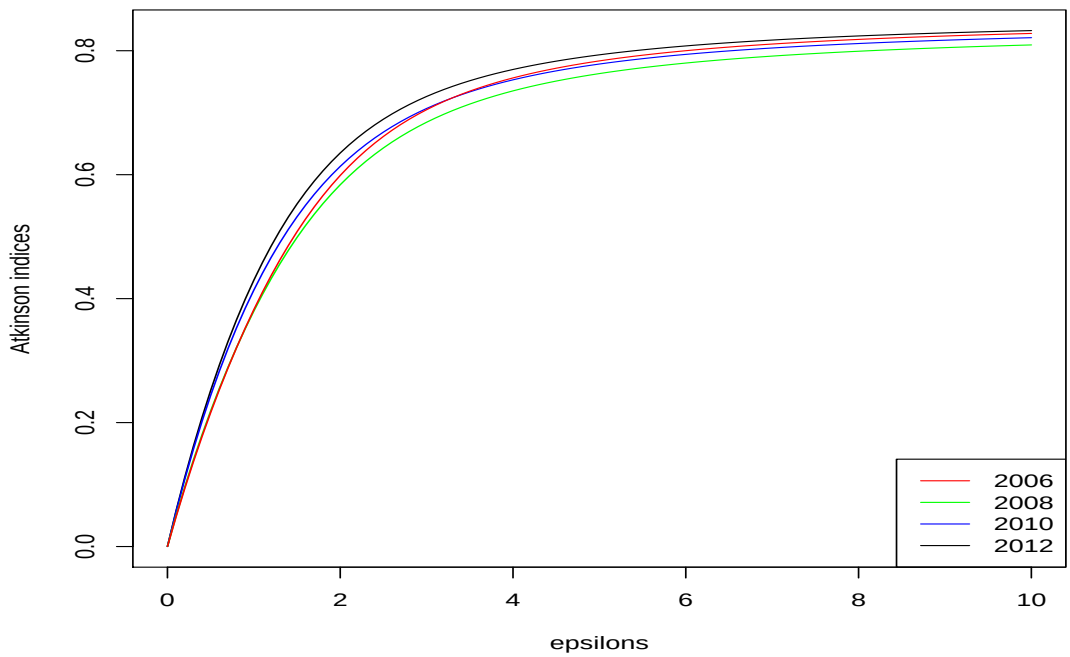


Figure 23: Atkinson indices with varying epsilon values for total wealth in the UK from 2006-2012.

From the plot it seems that the green line representing the 2008 situation has a similar Atkinson index to the 2006 situation, represented by the red line when epsilon is below one. As epsilon increases, the 2008 distribution diverges slightly and remains consistently below the others for the rest of the plot. This indicates that the 2008 situation was the most favourable state for the very poor, despite the fact that the results for first order dominance are not coherent with this. However, this result is consistent with previous results such as the Gini coefficient and the Lorenz curves - the distributions have the same ordering for these three.

15.5 Transition matrices

In this subsection we aim to construct a transition matrix between wealth groups with a distance of two years. In (Jenkins, 2000) he does this for income groups in our previous data, the British Household Panel Survey (BHPS) from 1991 to 1996 at a distance of

one wave. If there are m class boundaries $0 < \zeta < \cdots < \zeta_{m-1} < \infty$ and $0 < \xi < \cdots < \xi_{m-1} < \infty$ corresponding to two variables x and y then each element p_{ij} of the transition matrix P can be written as:

$$p_{ij} = \frac{\mathcal{P}(\zeta_{i-1} \leq x \leq \zeta_i \text{ and } \xi_{j-1} \leq y \leq \xi_j)}{\mathcal{P}(\zeta_{i-1} \leq x \leq \zeta_i)} \tag{53}$$

where $\zeta_0 = \xi_0 = 0$ and $\zeta_m = \xi_m = \infty$ As an alternative we used the *markovchain* package in R which easily allows for handling Markov chains. We first transform our vectors of wealths corresponding to the respective years into vectors of classes. We then mainly make use of the *markovchainFit()* function which produces a transition matrix directly, as well as a matrix of standard errors and upper and lower bounds. Additionally, a confidence level can be specified by the user with the default being 95% which was used here. The method used to calculate the probabilities can also be specified, with the default being a maximum-likelihood method. There is also the option for a Bayesian, Bootstrap or Laplace method. The maximum likelihood estimator of each element of the transition matrix P is:

$$\hat{P} = [\hat{p_{ij}}] = \left[\frac{n_{ij}}{ni}\right] \tag{54}$$

where $n_{ij}, i, j = 1, \ldots, k$ is the number of households initially in state i moving to state j in the next step. Additionally, $n_i = \sum_{j=1}^{k} n_{ij}$ is the initial number of households in state i and $n = \sum_{i=1}^{k} n_i$.

We use the same income classes as Jenkins in that they correspond to less than 0.5 times the mean, 0.5-0.75 times the mean, 0.75-1 times the mean and so on with the last group being over 1.5 times the mean or above. Due to the fact that we go from step $i-1$ to i, we can produce three matrices corresponding to 2006-2008 to 2008-2010, 2008-2010 to 2010-2012 and 2010-2012 to 2012-2014. We can then produce a mean transition matrix whereby the probabilities correspond to being in one of the groups on the left at state i and moving to another on the right at state j.

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
<0.5	0.83	0.09	0.04	0.02	0.01	0.01
0.5 - 0.75	0.35	0.24	0.19	0.08	0.05	0.09
0.75 - 1.0	0.21	0.22	0.21	0.10	0.07	0.19
1.0 - 1.25	0.14	0.15	0.21	0.12	0.09	0.28
1.25 - 1.5	0.10	0.11	0.18	0.12	0.09	0.38
>1.5	0.05	0.07	0.10	0.09	0.08	0.61

Table 6: Mean transition matrix for wealth groups from 2006-2014

As seen in the matrix, the probability of remaining in the lowest group is 0.83 - testifying to the fact that it is hard to escape poverty. It seems as though it is also likely for those in the top group to stay in this group as expected at a probability of 0.61 although

not as likely as staying poor. Those who jump from the lower groups to the last are most likely the recipients of inheritance or large gifts. Another explanation could be self-recruiting in executive positions by the most wealthy.

There is a steady increase in the probability of moving into the next subsequent group as we go through each group, testifying to the cumulative effect that wealth produces. Interestingly, there is a small chance (5%) that a household in the top group falls to the bottom which is at a much higher probability than expected. It is also interesting that there is a reasonable chance at a probability of 0.35 of moving from the second lowest group to the lowest, once again indicating that poverty can be hard to escape.

Note that the transition matrix is irreducible in that all states can be reached from every other state. It is also aperiodic in that it has period one. This is obtained from realising that the greatest common divisor of $\{k \geq 1 : p_{i,i}^k > 0\}$ is equal to one. This means that we can go on to produce an equilibrium distribution. A matrix containing standard errors of our previous probabilities can also be produced which is given:

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
<0.5	0.00450	0.00144	0.00101	0.00061	0.00042	0.00051
0.5-0.75	0.00562	0.00477	0.00419	0.00282	0.00211	0.00296
0.75-1	0.00466	0.00444	0.00451	0.00317	0.00275	0.00442
1-1.25	0.00451	0.00503	0.00597	0.00458	0.00390	0.00694
1.25-1.5	0.00502	0.00516	0.00651	0.00535	0.00479	0.00961
>1.5	0.00163	0.00187	0.00234	0.00222	0.00211	0.00580

Table 7: Mean standard error transition matrix for wealth groups from 2006-2014

The standard errors are very low for each p_{ij} which is an indication of a good model. The equilibrium distribution of P which is given by: $\pi = \pi P$ where π is the stationary distribution. This represents the long-term distribution of the probabilities of being in a specific group. We again turn to the *markovchain* package, this time using the *steadyStates* function.

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
Equilibrium	0.50723	0.12064	0.10410	0.05709	0.04021	0.17074
Actual	0.47194	0.11257	0.11444	0.05961	0.05063	0.19081

Table 8: Equilibrium and actual distributions of wealth groups

Clearly the equilibrium distribution is close to that of the actual distribution, with the middle four classes of the equilibrium distribution being very close to that of the true distribution. Where the most disparity occurs is at the bottom and top classes, with the proportion of very poor being over-estimated and the proportion of very wealthy being

under-estimated. A Kolmogorov–Smirnov test gives a statistic of 0.17 with a p-value of 1 meaning that we can conclude that there is no significant difference in the two distributions.

Prais (1955) was the first to study social mobility using a Markov model according to (Bartholomew, 1982). He also proposed a mobility index given by:

$$M_P = \frac{k - tr(P)}{k - 1} \tag{55}$$

where k is dimension of the transition matrix ($k = 6$ in our case). This index can be seen as an extension of the formula for the mean exit time of each group which is given by $\frac{1}{1 - p_{ii}}$ hence M_P can be seen as the harmonic mean of the mean exit times which is normalised by $k/(k-1)$. A situation of perfect immobility arises when there is no possibility to move into any other group which is of course when P is the identity matrix, I. This would be when $M_{P=I} = 0$.

In (Shorrocks, 1978) he attempts to provide various axioms for mobility indices. These are:

- Normalisation: $0 \leq M(P) \leq 1$
- Monotonicity: $P \succcurlyeq P' \rightarrow M(P) > M(P')$
- Immobility: $M(I) = 0$ iff $P = I$
- Perfect mobility: $M(P) = 1$ if $P = i'x$ with $x'i = 1$.

Where $P \succcurlyeq P'$ when P is more mobile than P' where P' and P are two distinct transition matrices. Prais' index only fulfils the immobility and monotonicity indices.

As we originally had three transition matrices, we can find M_P for each time period to see the evolution of social mobility for our scenario. This is given in the table below where we can see that Prais' index increases over the periods, indicating an increase in mobility. This is odd considering we would expect the opposite in accordance with the increase in inequality over the periods. The main reason for this is due to the fact that in the middle groups, the probabilities of remaining in the same group decreased - a dispersion of the middle class. Keep in mind that the Prais' index only takes into account the diagonal values of the matrix and so is not concerned with transitions from one group to another, different group. This highlights a potential flaw in the way that Prais defines mobility and so we explore various other indices that may capture a more realistic picture of inequality while also reducing dependency on one index.

The other two indices are both from (Bartholomew, 1982). The first one is $M(P) = k - \sum_i \pi_i p_{ii}$ and the second is $M(P) = \frac{1}{k-1} \sum_i \pi_i \sum_j p_{ij}|i-j|$.

Index	2006-2008	2008-2010	2010-2012	Mean
$M_P = \frac{k - tr(P)}{k-1}$	0.74	0.79	0.81	0.78
$M_P = k - \sum_i \pi_i p_{ii}$	0.54	0.60	0.68	0.59
$M_P = \frac{1}{k-1} \sum_i \pi_i \sum_j p_{ij}\lvert i-j\rvert$	0.86	0.76	0.68	0.8

Table 9: Mobility indices for different states

To avoid confusion, the mean value above does not correspond to the mean of the other three but the index generated from the mean transition matrix, although they give a very close result. Bartholomew's first index is coherent with Prais' index in that mobility rises over time due to the fact that only the diagonal is taken into account but his second index actually shows a fall in mobility as we would expect. This is because the non-diagonal elements are finally taken into account as well as the equilibrium distribution, finally giving a result we would expect.

The work above corresponds to total aggregate wealth. We also have access to the data of the components that make up this total wealth including physical, financial and property wealth. We can therefore produce additional transition matrices for these wealth components to assess mobility in the same way we have done before. These are given below:

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
<0.5	0.62	0.09	0.06	0.05	0.03	0.15
0.5-0.75	0.59	0.10	0.11	0.08	0.02	0.11
0.75-1	0.42	0.17	0.10	0.08	0.09	0.13
1-1.25	0.42	0.20	0.13	0.06	0.09	0.08
1.25-1.5	0.44	0.09	0.18	0.19	0.10	0.01
>1.5	0.30	0.06	0.05	0.04	0.03	0.52

Table 10: Mean transition matrix for physical wealth groups from 2006-2014

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
Equilibrium	0.51513	0.09885	0.07243	0.06104	0.03968	0.21288
Actual	0.49546	0.11438	0.07624	0.07139	0.03801	0.20453

Table 11: Equilibrium and actual distributions of physical wealth groups

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
<0.5	0.82	0.04	0.03	0.02	0.01	0.08
0.5 - 0.75	0.41	0.13	0.09	0.05	0.04	0.28
0.75-1	0.37	0.11	0.11	0.06	0.04	0.32
1-1.25	0.30	0.10	0.09	0.06	0.05	0.40
1.25-1.5	0.27	0.08	0.09	0.07	0.06	0.43
>1.5	0.25	0.09	0.08	0.06	0.05	0.47

Table 12: Mean transition matrix for financial wealth groups from 2006-2014

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
Equilibrium	0.62807	0.06371	0.05321	0.03246	0.02395	0.19859
Actual	0.60804	0.07081	0.05604	0.03151	0.02957	0.20402

Table 13: Equilibrium and actual distributions of financial wealth groups

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
<0.5	0.79	0.03	0.01	0.01	0.02	0.13
0.5-0.75	0.43	0.12	0.06	0.08	0.06	0.25
0.75-1	0.38	0.04	0.12	0.07	0.11	0.28
1-1.25	0.35	0.15	0.04	0.00	0.03	0.43
1.25-1.5	0.40	0.06	0.07	0.11	0.08	0.28
>1.5	0.34	0.05	0.05	0.06	0.04	0.45

Table 14: Mean transition matrix for property wealth groups from 2006-2014

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
Equilibrium	0.63455	0.04657	0.02864	0.03257	0.02972	0.22795
Actual	0.64089	0.04421	0.03076	0.02898	0.03313	0.22201

Table 15: Equilibrium and actual distributions of property wealth groups

Despite being the components that make up total wealth, the transition matrices above are very different to the original one concerning total wealth. Most notably, the probabilities of remaining in the top group are much lower for these types of wealth at 0.52, 0.47 and 0.45 compared to total wealth at 0.61. It also seems as though it is much easier for those in the higher groups to end up in the lowest group in these three later matrices when compared to the original. Looking at the matrix corresponding to physical wealth, it is much easier to escape poverty, with a surprising probability of 0.15 of moving from the lowest to the highest group, reflected by its equilibrium distribution. A similar result occurs for property wealth at a probability of 0.13. The most reasonable explanation for this is that a large portion of inheritance tends to be in property and physical assets

rather than in financial assets.

The equilibrium distributions for physical, property and financial wealth are similar to their respective actual distributions which is a good sign that we still are producing good models even if they are not coherent with the original. Again, the differences in the actual and equilibrium distributions are largest in the bottom and top wealth group as seen before, with the middle groups having very similar proportions.
KS tests gives the same statistic of 0.17 for physical, property and financial wealth again alluding to no significant difference between the two distributions for each type of wealth.

We can also provide analysis regarding the mobility for these types of wealths where the index given corresponds to the mean transition matrix for each type of wealth. The indices for total wealth are also included for comparison without having to refer to the previous table.

Index	Property	Financial	Physical	Total
$M_P = \dfrac{k - tr(P)}{k-1}$	0.89	0.87	0.90	0.78
$M_P = k - \sum_i \pi_i p_{ii}$	0.62	0.63	0.45	0.59
$M_P = \dfrac{1}{k-1} \sum_i \pi_i \sum_j p_{ij} \lvert i-j \rvert$	1.31	1.10	1.54	0.8

Table 16: Mobility indices for various types of wealths

This result reinforces the claim that it is easier to move between wealth groups when considering physical or property wealth when compared to total or financial wealth for the first index. However, physical wealth is much less mobile than the others when considering Bartholomew's first index which is surprising. For Bartholomew's second index, we can see that mobility is much higher for the three components than total wealth, giving further evidence to the fact that there must be other types of wealth not available in the data given with much a much lower mobility as to explain the difference.

We can also consider creating our wealth groups based on the median rather than the mean as performed previously. This allows us to see how the definition of poverty affects the transition probabilities. We only consider total wealth.

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
<0.5	0.80	0.12	0.02	0.02	0.02	0.02
0.5-0.75	0.33	0.27	0.03	0.14	0.09	0.14
0.75-1	0.31	0.24	0.03	0.13	0.10	0.19
1-1.25	0.09	0.18	0.02	0.17	0.17	0.36
1.25-1.5	0.09	0.14	0.01	0.17	0.16	0.43
>1.5	0.02	0.05	0.00	0.07	0.09	0.76

Table 17: Mean transition matrix for total wealth groups when considering the median

It seems as though the probability of staying in the lowest group is lower when considering the median rather than the mean, with it also being clear that it is much easier for high amounts on wealth to be retained. This is seen by a probability of 0.76 to remain in the highest wealth group for the median, contrasted by a figure of 0.61 for the mean. This is also supported by the fact that the probability of transitioning from the highest to the lowest group is lower when considering the median rather than the mean.
It seems as though overall it is easier for transitions into higher wealth groups when concerned with the median. This is most likely due to the fact that the medians for our concerned distributions are all considerably lower than their respective means, resulting in many more households lying in higher groups for this implementation than compared to when using the mean. This might be offering a slightly biased approach and so we conclude that the mean is more suitable in this regard. This is reinforced by the table of actual and equilibrium distributions below where we can see that nearly 40% of households lie in the top group compared to 20% for the mean.

Wealth group	<0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	>1.5
Equilibrium	0.33574	0.11946	0.01395	0.0799	0.08	0.37095
Actual	0.30932	0.11771	0.03402	0.09081	0.07360	0.37453

Table 18: Equilibrium and actual distributions of total wealth groups when considering the median

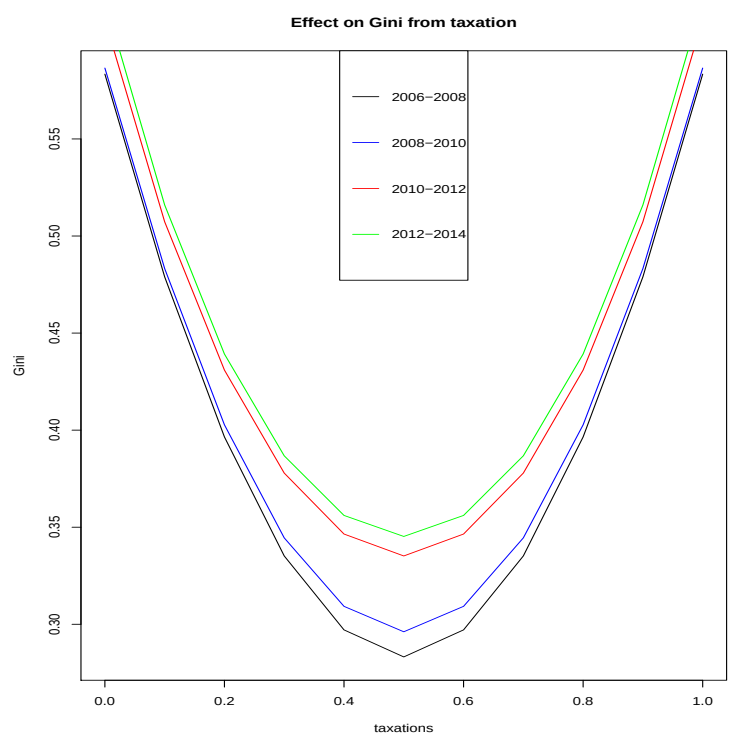
15.6 Taxation revisited

Additionally, we can perform the taxation simulations outlined previously in Section 12 for this data. We will perform simulations in the same order as before i.e flat taxation, quantile taxation and then UK-style taxation. The motivations for doing this is that we already have the functions produced earlier and so the results can be obtained easily. We also expect more realistic results compared to when we used the Mittag-Leffler distribution due to the fact that this data is of course more realistic. We produce the table below where the numbers on the top correspond to taxation proportions (i.e 0.5 = 50% taxation). A flat taxation rate of 0 of course refers to no taxation present and so these indices are the original Gini values.

Years	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
2006-2008	0.58	0.48	0.40	0.34	0.30	0.28	0.30	0.34	0.40	0.48	0.58
2008-2010	0.59	0.48	0.40	0.34	0.31	0.30	0.31	0.34	0.40	0.48	0.59
2010-2012	0.61	0.51	0.43	0.38	0.35	0.34	0.35	0.38	0.43	0.51	0.61
2012-2014	0.62	0.52	0.44	0.39	0.36	0.35	0.36	0.39	0.44	0.52	0.62

Table 19: Effect on the Gini coefficient of each time period through flat taxation

The table shows the same as the figure below, where we can see that a U-shaped curve is produced for each time period. This is expected due to the fact that a U-shaped curve was also present when instead using Mittag-Leffler random variables. Each subsequent time period has a curve further along the y-axis due to the fact that inequality increased over these periods. Note that when using Mittag-Leffler random variables as done before, we only achieved a reduction in the Gini of around 0.08 but in this case there is a reduction of 0.3 in the Gini for 2006-2008 at a rate of 50% - a much larger difference. It can also be seen that the maximum reduction in inequality achieved for each time period decreases over time. For example if we keep looking at the 50% rate there is a reduction of 0.29 in 2008-2010 and 0.27 in 2010-2012 and 2012-2014. Therefore we can interpret our previous simulations with Mittag-Leffler variables as an extreme possible convergence past the 2012-2014 period.

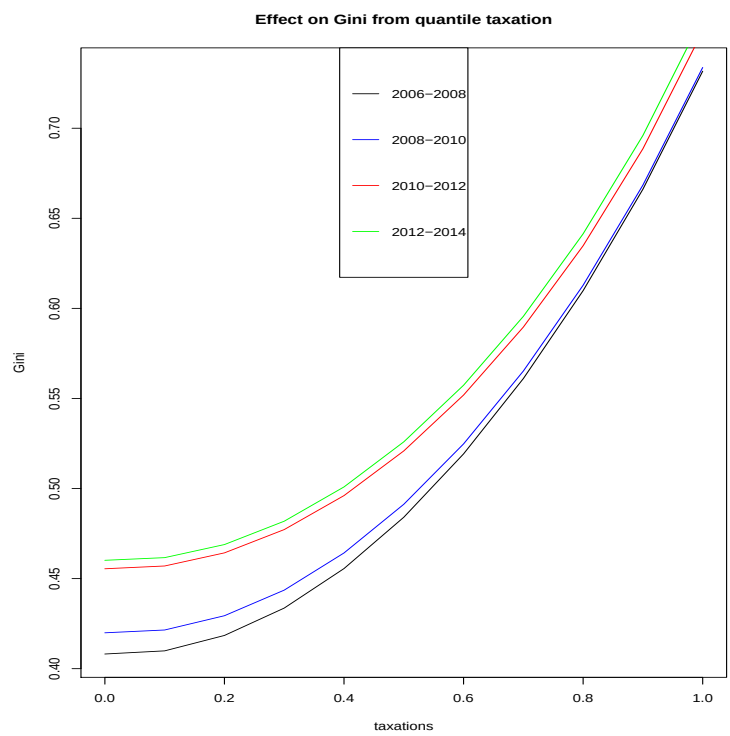


We now investigate quantile taxation to see if it produces a higher reduction in inequality for the concerned time periods. As a reminder, the taxation rates in the table below represent the level of leeway applied. For example, if a leeway of 0.2 is provided, a household in the $90th$ percentile would be taxed 70%.

Years	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
2006-2008	0.41	0.41	0.42	0.43	0.46	0.48	0.52	0.56	0.61	0.67	0.73
2008-2010	0.42	0.42	0.43	0.44	0.46	0.49	0.52	0.57	0.61	0.67	0.73
2010-2012	0.46	0.46	0.46	0.48	0.50	0.52	0.55	0.59	0.63	0.69	0.75
2012-2014	0.46	0.46	0.47	0.48	0.50	0.53	0.56	0.60	0.64	0.70	0.76

Table 20: Effect on the Gini coefficient of each time period through quantile taxation

We can see that the reduction in the Gini is not as profound as when performing flat taxation, with the largest reductions corresponding to no leeway with reductions of 0.17 for the first two periods, 0.15 for 2010-2012 and 0.14 for 2012-2014. Surprisingly, this is a different result as to when using Mittag-Leffler data where we saw that the reduction in inequality was almost the same for all three methods. The graph below provides a visual representation:



In both of the graphs, we can see a jump of the lines from 2008-2010 to 2010-2012 which most likely corresponds to the increase in inequality during these periods. Now we move onto UK-style taxation:

Years	Original Gini	Normal UK-taxation	Random reallocation
2006-2008	0.58	0.39	0.50
2008-2010	0.59	0.40	0.50
2010-2012	0.61	0.43	0.53
2012-2014	0.62	0.44	0.53

It seems as though normal UK-style taxation provides a greater reduction in the Gini when compared to quantile taxation but less when compared to flat taxation, meaning we can conclude that flat taxation is the current preferred method. It was attempted to randomly allocate the taxation back to the population instead of to the poorest as done in the previous section with Mittag-Leffler data. However, in this case this was repeated 100 times for each time period to produce a more solid result but there was not much change when compared to just performing reallocation once and so this is not included. We now provide four Lorenz curves corresponding to each time period with multiple lines depicting the new curves produced after the types of taxation.

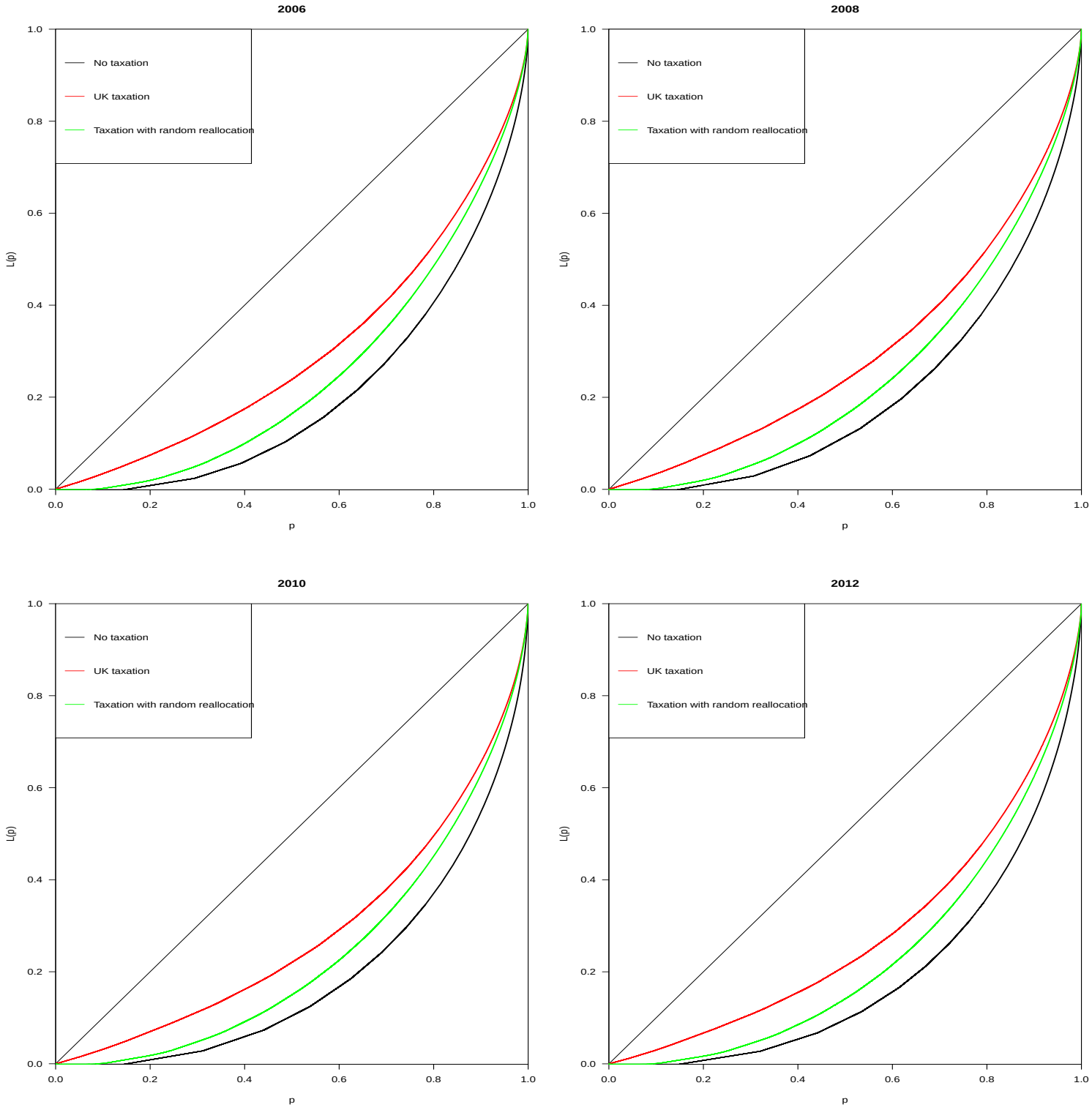


Figure 24: Lorenz curves for taxation

The changes in the curves for each time period are relatively consistent, with the same shapes being produced for each of the periods. In every case it is that random reallocation provides a slight decrease in the Gini, but a more profound reduction is produced by redistribution directly to the poor.

16 Conclusion

In this dissertation we have managed to prove our concerned conjecture, and we have achieved some good results with the empirical data. The section on transition matrices is likely the most profound result with the main transition matrix for total wealth offering a realistic interpretation of social mobility.

To improve the dissertation we could have performed some data analysis with particular relevance to Lorenz curves for different countries in addition to the UK. This would have allowed for a comparison between inequalities in various global societies. In particular it would have been interesting to see the global trend of overall inequality. For example, we could have produced transition matrices for multiple countries in order to compare social mobility with particular focus on the difficulty of escaping poverty. However, the second set of data that was used from the Office of National Statistics (ONS) was very comprehensive which may not be the case for other countries, especially those with less concern for data collection. This would make comparisons very difficult due to the potential inconsistencies that would be present.

We could have also aimed to prove convergence of the sequence of empirical concentration curves with infinite mean to perfect inequality. Goldie does this for finite mean in addition to Lorenz curves and so applying the same methodology as we did with the proof would potentially attain the result. Most likely convergence would be the same, with uniform convergence not being attained and instead having convergence under the J_1 metric.

Another potential improvement would be another Bayesian implementation for a distribution besides the Gamma whereby it would provide a better fit for the data in that the standard error would be lower. This could have been done with any of the frequently used distributions to model incomes or wealth such as the Singh-Maddala, Dagum or any of the variations of the Pareto. In addition it would have been interesting to examine second-order stochastic dominance more, with a potential formulation of a t-test in the same way that Davidson and Duclos did for first-order dominance.

Some more information regarding the discussions and figures that Piketty produces in his book may have also been a good idea. In particular, the reasons for the changes in inequality over the years and the ethics that underline them could have been explored more.

17 References

1. Lorenz, M., 1905. Publications of the American Statistical Association, 9(65-72), p.218

2. Gastwirth, J., 1971. A General Definition of the Lorenz Curve. Econometrica, 39(6), p.1037.
3. Damgaard, C. and Weiner, J., 2000. Describing inequality in plant size of fecundity . Ecology, 81(4), pp.1139-1142
4. Gastwirth, J., 1972. The Estimation of the Lorenz Curve and Gini Index. The Review of Economics and Statistics, 54(3), p.306.
5. Lamb, E., 2012. Ask Gini: How To Measure Inequality. Scientific American. Available at: https://www.scientificamerican.com/article/ask-gini/ [Accessed 13 August 2020].
6. Basu, K., 2018. The Ethics Of Reducing Inequality. Brookings. Available at: https://www.brookings.edu/opinions/the-ethics-of-reducing-inequality/ [Accessed 6 June 2020]
7. Fulger, D., Scalas, E. and Germano, G., 2008. Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation. Physical Review E, 77(2).
8. Shorrocks, A., 1983. Ranking Income Distributions. Economica, 50(197), p.3.
9. Duclos, J. and Araar, A., 2006. Poverty And Equity. New York, NY: Springer.
10. Atkinson, A., 1970. On the measurement of inequality. Journal of Economic Theory, 2(3), pp.244-263.
11. Neves Costa, R. and Pérez-Duarte, S., 2020. Not All Inequality Measures Were Created Equal. [online] Ecb.europa.eu. Available at: https://www.ecb.europa.eu/pub/pdf/scpsps/ecb.sps31~269c917f9f.en.pdf [Accessed 13 August 2020].
12. Scalas, E., Gorenflo, R. and Mainardi, F., 2004. Uncoupled continuous-time random walks: Solution and limiting behavior of the master equation. Physical Review E, 69(1), p.8.
13. Wash, R., 2008. [online] www.rickwash.com. Available at: http://www.rickwash.com/papers/cscw08-appendix/powerlaw.R [Accessed 13 August 2020].
14. Iser.essex.ac.uk. n.d. British Household Panel Survey (BHPS) - Institute For Social And Economic Research (ISER). [online] Available at: https://www.iser.essex.ac.uk/bhps [Accessed 13 August 2020].
15. ineq: Measuring Inequality, Concentration, and Poverty, Achim Zeileis, 2014, https://CRAN.R-project.org/package=ineq
16. mixdist: Finite Mixture Distribution Models, Peter Macdonald and with contributions from Juan Du, 2018, https://CRAN.R-project.org/package=mixdist
17. P. Minka, T., 2002. Estimating A Gamma Distribution. [online] Tminka.github.io. Available at: https://tminka.github.io/papers/minka-gamma.pdf [Accessed 13 August 2020].
18. Ons.gov.uk. 2019. Total Wealth: Wealth In Great Britain - Office For National Statistics. [online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/totalwealthwealthingrea [Accessed 13 August 2020].
19. Chamberlain, E., 2015. Wealth In Great Britain Wave 4: 2012 To 2014 - Office For National Statistics. [online] Ons.gov.uk. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/compendium/wealthingreatbritainwave4/2012to2014/wealthingreatbritainwave42012to2014 [Accessed 13 August 2020].
20. Foster, J., Greer, J. and Thorbecke, E., 1984. A Class of Decomposable Poverty Measures. Econometrica, 52(3), p.761.
21. Davidson, R. and Duclos, J., 2013. Testing for Restricted Stochastic Dominance. Econometric Reviews, 32(1), pp.84-125.
22. Jenkins, S. P. (2000). Modelling household income dynamics. Journal of Population Economics, 13:529–567.
23. Discrete Time Markov Chains with R, Giorgio Alfredo Spedicato, 2017, https://journal.r-project.org/archive/2017/RJ-2017-036/index.html
24. Prais, S., 1955. The Formal Theory of Social Mobility. Population Studies, 9(1), p.72.
25. Bartholomew, D. (1982). Stochastic Models for Social Processes. Wiley, London, 3rd edition
26. Shorrocks, A. F. (1978). The measurement of mobility. Econometrica, 46(5):1013–1024
27. https://en.wikipedia.org/wiki/Lorenz_curve [Accessed 6 June 2020]
28. Chung, K., 1968. A Course In Probability Theory. New York: Harcourt, Brace and World.