Theory and Evidence on

Estimating Global Income Inequality from Limited National Data Points between 1990 and 2010

By

Abhinav Mittal

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science

Undergraduate College

Leonard N. Stern School of Business

New York University

May 2016

Professor Marti G. Subrahmanyam

Faculty Adviser

Professor Rangarajan K. Sundaram

Thesis Adviser

ABSTRACT

An accurate global Gini coefficient calculation requires an accurate global income distribution. In certain cases – particularly where direct national income distribution data is scarce – it is useful to indirectly estimate the global income distribution from more readily available national statistics. This paper puts forward a model to calculate the global Gini coefficient from national Gini coefficients and GDP per capita in three years: 1990, 2000 and 2010. The model is applied to a sample of 35 countries covering 76% of the global population, for which data was available at the World Bank database. The model finds that the global Gini coefficient for 1990, 2000 and 2010 are 0.44, 0.44 and 0.46 respectively. This suggests that global income inequality remained virtually the same between 1990 and 2000 and increased slightly between 2000 and 2010. The model may be applied more generally to calculate the Gini coefficient of any group for which the Gini coefficient, GDP per capita and population of sub-groups is known.

INTRODUCTION

Income inequality has come to the fore as one of the most important socio-economic topics that dominates discussion in American politics and media. Over the last few years, a book on the subject by an economist – *Capital* by Thomas Picketty – became a New York Times bestseller and US President Barack Obama noted income inequality as the "defining challenge of our time"[1].

The anxiety and debate over income inequality is motivated by the widening gap between the rich and the poor in developed countries over the last three decades. As of December 2014, the richest 10% of the population in OECD countries earned 9.5 times more than the poorest 10%[2]. There have been a number of theories put forward for the increase in income inequality; two of the most widely discussed are increased automation and increased global economic integration. Proponents of the latter theory argue that as the world became more inter-connected, low-skill jobs that were done by low-income individuals in a developed country moved to a developing country through international trade, outsourcing or other means. One may then expect to see an increase in income inequality in the developed country, but also a decline in income inequality between the developed and the developing country. There is significant interest to investigate whether this is the case, as

---

1 The White House: Remarks by the President on Economic Mobility (2013).
2 OECD: Society at a Glance (2014).

seen by an increase in the number of research papers exploring changes in income inequality[3]. Yet such research is often held back due to dearth of accurate data, and differences and opaqueness in methodology employed by researchers to calculate global income inequality[4].

Global income inequality may be defined in a number of ways. Milanovic provides a useful classification[5]: type one inequality i.e. inequality among countries in their levels of average per capita income; type two inequality i.e. between-country inequality, which measures inequality between individuals where each person is assigned the per capita income of the country he or she resides in; and type three inequality i.e. global interpersonal inequality, where each individual is assigned his or her own income. The choice of inequality to be studied depends on the motivation behind the research. This paper is concerned with type-three inequality for intrinsic reasons: a vast difference in incomes and living standards among individuals around the world may be considered unjust, and must be investigated.

LITERATURE REVIEW

The most widely used measure of income inequality is the Gini coefficient. When people in a group (for instance, a nation) are arranged from poorest to richest on the x-axis and the corresponding cumulative income shares are plotted on the y-axis, the Gini coefficient measures the distance between the line of equality and the cumulative income function (also known as the *Lorenz Curve*), as a fraction of the area under the line of equality. A Gini coefficient of 0 indicates perfect equality, while a Gini coefficient of 1 indicates that one person owns all the wealth in the group. The Gini coefficient is usually calculated and compared nationally because surveys that collect data on income are conducted on a national level. This allows one to create a national income distribution from which the national Gini coefficient may be calculated. A global Gini coefficient however, would require a global income distribution, which is not directly available- there is no body that conducts a global survey of incomes. As a result, researchers must employ other ways to estimate a global Gini coefficient.

While it is theoretically possible to combine the results of national surveys to create a global distribution, in practice, this proves to be problematic. The only major researcher who uses direct survey data to

[3] Sudhir Anand and Paul Segal, What Do We Know Abut Global Income Inequality? (Journal of Economic Literature: 2008).
[4] Sudhir Anand and Paul Segal, What Do We Know Abut Global Income Inequality? (Journal of Economic Literature: 2008)
[5] Branko Milanovic, Global Income Inequality By the Numbers: In History and Now– An Overview: 2012)

create a global income distribution is Milanovic.[6] While the current direct global income distribution data that he works with spans the vast majority of the world and uniquely allows for a 'true' or direct global distribution, it is limited in time: data is available for certain 5 year intervals such as 1988, 1993, 1998 and thereafter, thus forcing the use of indirect methods to compare inequality to time periods prior to the availability of direct income data. In *What Do We Know about Global Income Inequality?*, Anand and Segal note that "typically, the papers that estimate global inequality do not use primary data from household surveys, but rather compilations of secondary data."[7] Even secondary data on national income distribution is not always available. China, for instance, only began to conduct and make available national income surveys in the 1980s; secondary data on income inequality in the country prior to that is less reliable. (Even today, national income distribution data on several countries, particularly in Africa, is still not available.)

The lack of data is not the only challenge. There is also no widely accepted methodology on how secondary data should be used to estimate the global income distribution. These two factors have resulted in little consensus over whether global income inequality is trending up or trending down. A comparison of various researchers' estimates of Gini coefficients over time compiled by Anand and Segal is provided as Figure A in the Appendix. The figure shows that while there is sizable variation between results on trends, most research suggests global income inequality is quite high, with a Gini coefficient ranging from 0.6 to 0.7. That puts global interpersonal inequality at par with some of the most unequal countries in the world, such as Namibia (0.61)[8].

A number of different techniques have been expounded to extrapolate the global income distribution from limited national data. Some researchers (such as Bourgigon and Morrison in *Inequality Among World Citizens: 1820 – 1992*[9], and Sala-i-Martin in *The World Distribution of Income: Falling Poverty and … Convergence, Period*[10]) use decile national income shares, assuming that incomes are equally distributed within

---

[6] The World Bank: Branko Milanovic, True World Income Distribution, 1988 and 1993: First calculations based on household surveys alone: 1999
[7] Anand and Segal, What Do We Know about Global Income Inequality? (Journal of Economic Literature: 2008).
[8] Gini Index, World Bank estimate
[9] Francois Bourguignon and Christian Morrison, Inequality among World Citizens: 1820 – 1992 (The American Economic Review: 2002)
[10] Xavier Sala-i-Martin, The World Distribution of Income: Falling Poverty and … Convergence, Period (Quarterly Journal of Economics: 2006)

each decile, and aggregate national data to get a global income distribution. While such an assumption of within-fractile equality leads to a downward bias in income inequality, subsequent research by Dowrick and Akmal in *Contradictory Trends in Global Income Inequality: A Tale of Two Biases*[11], suggests that grouping by deciles may account for 95 percent of the actual value.

In scenarios where one can not obtain fractile data, one must contend with the limited data points available. For instance, where survey data was not available for certain countries, Sala-i-Martin used data from neighboring countries to estimate a country's Gini coefficient[12]. It may then be useful to create a national income distribution based on only a few widely available data points such as GDP per capita and the Gini coefficient (here estimated through indirect methods). Chotikapanich, Valenzuela and Rao put forward a model to do so as they estimated national (and subsequently global) income distribution from national Gini coefficient and GDP per capita data in *Global and Regional Inequality in the Distribution of Income: Estimation with Limited and Incomplete Data*[13]. They assume that income distribution in each country follows a log-normal pattern, based on which they plot the income of the residents of the country. They then create an aggregate cumulative income distribution to calculate the global Gini coefficient. In their paper, however, the authors concede that there is little evidence to support the assumption of income being distributed log-normally and, I am not aware of subsequent work that either proves or disproves the assumption, which calls into question the validity of the resulting Gini coefficient.

This paper builds on such efforts to calculate the Gini coefficient from limited data points. In this paper, I suggest a model to create a global income distribution based on the Gini coefficients, GDP per capita and population of countries without assuming log-normality. The latter two metrics are widely available, and estimates of the first are available at the World Bank database going back further than from when a direct global income distribution may be calculated. Such a model may be employed to calculate the income distribution of a number of countries whose Gini coefficients have been indirectly estimated. This is useful for including those

[11] Steve Dowrick and Muhammad Akmal, Contradictory Trends in Global Income Inequality: A Tale of Two Biases (Australian Bureau of Agriculture and Resource Economics: 2005)

[12] Xavier Sala-i-Martin, The World Distribution of Income: Falling Poverty and … Convergence, Period (Quarterly Journal of Economics: 2006)

13 Duangkamon Chotikapanich, Rebecca Valenzuela and Prasad D.S. Rao, Global and Regional Inequality in the Distribution of Income: Estimation with Limited and Incomplete Data (1997)

countries in the global analysis on whom survey data still does not exist, but perhaps more importantly, it is useful to analyze global income inequality in decades past, where income distribution was missing for a large section of the world. The resulting global distribution may help researchers compare trends in global income inequality.

MODEL

I estimate the group global income inequality using national Gini coefficients, GDP per capita and Population data by assuming that the cumulative income function (the *Lorenz curve*) follows the form $F(p) = Y p^{1+\alpha}$, where p is the population percentile, Y is the GDP of the country and α is positive.

Since the Gini coefficient is a measure of the distance between the equality line and the Lorenz curve as a proportion of the area under the Lorenz curve, the Gini coefficient should indicate the implied area under the Lorenz curve. Based on the area under the curve, one can obtain the implied α of the cumulative income function $F(p) = Y p^{1+\alpha}$ for that country. One may then calculate the number of people between any two income levels and get a national distribution of income. The aggregation of national incomes between defined income values gives the global income distribution, from which one could approximate α for the global cumulative income distribution function $F(p) = Y p^{1+\alpha}$. Using the same intuition used to convert a national Gini coefficient to a value of α, one could use the global α to calculate the global Gini coefficient.

The model is formally laid out as follows:

Gini Coefficient Model

May 12, 2016

1 Country

For a given country, let the GDP be denoted Y and the population be denoted Q. We make an assumption that the cumulative income by percentile can be denoted by the function

$$F(p) = Yp^{1+\alpha},$$

where p is the population percentile and α is positive. That is, the sum of the income of all people at or below the pth percentile is $F(p)$. α must be positive since $F(p)$ is a convex function. By definition, the poorest x% of people can not have more than x% of income.

The percentile is calculated by $p = \frac{x}{Q}$, where x is the number of people at or below the pth percentile. Hence

$$F(x) = Y\left(\frac{x}{Q}\right)^{1+\alpha}.$$

To convert the cumulative income distribution to a non-cumulative income distribution, we differentitate the function as follows:

$$f(x) = \frac{d}{dx}F(x) = \frac{Y}{Q}(1+\alpha)\left(\frac{x}{Q}\right)^{\alpha}.$$

Interpret this as the xth person, ranked by income, makes approximately $f(x)$ in income. Here, we substitute $y = \frac{Y}{Q}$, for GDP per capita. Therefore, to find the number of people making less than an income level L, you find the xth person who makes L; that is, solve for x in terms of L. Let

$$L = f(x) = \frac{Y}{Q}(1+\alpha)\left(\frac{x}{Q}\right)^{\alpha} \quad \Longrightarrow x = Q\left(\frac{L}{y(1+\alpha)}\right)^{1/\alpha}.$$

This gives the function $x(L)$, called the *population function*, of people making less than or equal to L. Finally, to find the number of people making between two income levels $L_1 < L_2$, subtract $x(L_2) - x(L_1)$.

2 Global

Let there be countries $C_1, \dots, C_n$. Let their corresponding population functions be denoted $x_1, \dots x_n$ respectively. Then, we can define a global population function

$$x_G(L) = \sum_{i=1}^{n} x_i(L),$$

which gives the number of people globally whose income is less than or equal to L.

2.1 Applying model to the global distribution

Let us impose that same model for global income distribution as above :

$$F_G(x) = Y_G \left(\frac{x}{Q_G}\right)^{1+\alpha},$$

where Y_G is global GDP, Q_G is global population and F_G gives the total income of the lowest x earners in the world. Then, the predicted population function globally, the inverse of F_G, will have the analogous form

$$\tilde{x}_G = Q_G \left(\frac{L}{y_G(1+\alpha_G)}\right)^{1/\alpha_G},$$

where y_G is the global GDP/capita and α_G is the global α parameter.

2.2 Approximating α

We approximate α_G by attempting to reconcile the population function data x_G based on the aggregated national income distributions with the model-predicted population function $\tilde{x}_G$. To get a better fit, we take the log of $\tilde{x}_G$,

$$\log \tilde{x}_G = \log Q_G - \frac{1}{\alpha} \log(y(1+\alpha_G)) + \frac{1}{\alpha} \log L.$$

Let there be 16 sample points $L_n = 2^n$ for $n = 1, 2, \dots, 16$. The sample points are chosen at multiples to achieve even spacing of sampling points under the log transform. Also, from the data at hand, it has been noted that no income was recorded higher than 2^{16} in 1990. For each country, we compute $x_G(L_n) = \sum_{i=1}^{n} x_i(L_n)$ and take the log of the computed x_G. On the other hand, for each L_n (take $\log L_n$), compute the predicted value $\log \tilde{x}_G(L_n)$ based on the model equation above. Then, compute the total square error,

$$E = \sum_{n=1}^{16} \left(\log x_G(L_n) - \log \tilde{x}_G(L_n)\right)^2.$$

Let α be chosen such that this error is minimized.

2.3 Computing the Gini coefficient

We return to the original model $F_G(p) = p^{1+\alpha_G}$, where p is the percentile. Here, there is no scaling by Y_G since we are not interested in the gross cumulative income, but the percent of total income. The Gini coefficient is given by

$$\Gamma = 2\int_0^1 p - F_G(p)\,dp = 2\int_0^1 p - p^{1+\alpha_G}\,dp = 2\left[\frac{p^2}{2} - \frac{p^{2+\alpha_G}}{2+\alpha_G}\bigg|_0^1\right] = 1 - \frac{2}{2+\alpha_G}.$$

DATA

I apply the model outlined above to a subset of the global population in three years: 1990, 2000 and 2010. The selection of countries for inclusion in the model was based on the availability of data.

I sought to collect the following metrics on each country: Gini coefficient (whether directly or indirectly estimated), GDP per capita and Population, and included all countries into my sample for whom the data was available in the World Bank database for the three relevant years. While GDP per capita and Population data is available for almost every country, Gini coefficient data is sparser. As a result, the model can only account for a portion of the world and does not adequately account for 'global' income inequality. In order for the model to better represent global data, I required that the 20 most populous countries must be present in the dataset. If the Gini coefficient for one of those countries was not listed for the relevant years, I estimated it by taking an average of the Gini coefficients for the two closest years, weighted by the gap to the relevant year (for instance, Brazil's 2000 Gini coefficient is estimated as an average between the 1999 and 2001 Gini coefficients). The sample thus produced accounted for 35 countries, representing 76% of the world population and 95% of the world income, as of 1990. I applied the model to this sample of countries. The expansion to a broader sample admittedly comes at the cost of some accuracy.

In addition, the World Bank lists several estimates of the Gini coefficient for each country in each year (as calculated by different organizations or researchers), along with a classification indicating the quality of the data. I selected the Gini coefficient of the highest quality in each entry. Once again, this is only possible at the cost of some consistency.

Since the goal of this study was to estimate the differences in living standards in the global population, it was imperative to consider incomes based on Purchasing Power Parity. Other researchers may choose to compare incomes based on market exchange rates depending on the motivation behind their study. To compare income inequality over the three-decade span, it was also necessary to account for inflation. As a result, I selected the World Bank data which provided PPP-adjusted income by country in current dollar terms.

The countries included in the model are listed in Figure B in the Appendix.

RESULTS AND DISCUSSION

The global Gini coefficient calculated over the three years surveyed is as follows:

Year	Gini Coefficient
1990	0.44
2000	0.44
2010	0.46

The model suggests that global income inequality stayed steady between 1990 and 2000, and increased slightly between 2000 and 2010.

However, there are several causes of concern that preclude me from making a definitive assertion to the effect. The biggest one is that the model encompasses only 35 countries, which account for 76% of the world's population and 95% of world income, as of 1990. Since some of the poorest countries were omitted from the sample, the estimated Gini coefficient will be an underestimation of global income inequality. Indeed, most research estimates global Gini coefficient in the last few decades to be between 0.6 and 0.7[1]. However, since the sample of countries over the three years is consistent, the change in the Gini coefficient does suggest an increase in income inequality in the sample countries between 2000 and 2010.

Secondly, the data on the World Bank database is a compilation of the work of several different researchers who have indirectly calculated the Gini coefficient using different methodologies. The inclusion of indirectly calculated Gini coefficients lends noise to the data. In addition, surveys themselves are not completely reliable sources of people's income. Deaton reported in 2005 that the rich tend to under report their income in surveys, while the poor are often inadvertently excluded from the survey population[2]. This further suggests that the true Gini coefficient would be higher than what the model predicts.

In addition to calculating the Gini coefficient for the three defined years, I conducted a test to measure the efficacy of the model. I plotted the Predicted Log X against the Log X to reflect how closely the distribution

[1] Anand and Segal, What Do We Know about Global Income Inequality? (Journal of Economic Literature: 2008).
[2] Anand and Segal mentioned in their article What Do We Know about Global Income Inequality
Angus Deaton, Measuring Poverty in a Growing World (or Measuring Growth in a Poor World. (Review of Economics and Statistics: 2005))

matches what the model predicts (X is a measure of how many people in the world are within a defined income bracket). In all three years – 1990, 2000 and 2010 – there is little error between the terms, reflecting that the model is a good predictor of the distribution. (The plots are included in the Appendix as Figures C, D and E). However, this must not be construed as evidence that the model would accurately calculate the global Gini coefficient. Instead, the plot shows that the aggregate global distribution as predicted by the model accurately reflects the national income distribution based on the factors included in the model. In addition, the closeness of fit also lends credence to the idea that the income distribution as expressed as a function substantially captures the data expressed in fractiles, but such an assertion would need to be evaluated further.

CONCLUSION AND FUTURE RESEARCH

The conclusion reached from this study is that there is evidence that income inequality in the sample countries stayed steady between 1990 and 2000, and increased from 2000 and 2010. However, the research is not definitive considering the concerns previously listed, and the subject needs to be evaluated further to get more accurate estimates. Perhaps the more important outcome of the research is that the model itself may be applied to calculate global income inequality more generally, pending validation described below.

To test whether the model may be used more broadly, the next validation step would be to apply it to a dataset where survey data on the income distribution of the broader group (world / nation) as well as the sub-groups (countries / states) is available, and the Gini coefficients of each sub-group that are input into the model are calculated directly from the sub-group income distribution distribution. Then, one would compare the Gini coefficient that the model would produce to the known group Gini coefficient. If the predicted Gini coefficient matches the known Gini coefficient, there would be evidence that the model is successful and could be applied to calculate the global Gini coefficient. I hope to conduct this analysis in a future study.

As data collection, reporting and analysis on income inequality becomes more sophisticated, researchers will be able to calculate global income inequality directly with greater confidence and accuracy. To compare how income inequality has changed over time, researchers may look towards past data to understand income distribution indirectly, for which models like this one may prove to be useful. Subsequent research in the field may help explain global income inequality as a function of other variables such as increased global economic integration, through trade, outsourcing and other mechanisms.

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Rangarajan Sundaram for his invaluable support at every step, from formulation of the idea to execution. I would also like to thank the Program Coordinator Marti Subrahmanyam for giving me the opportunity to take part in the Stern Honors Program and for his feedback and guidance throughout the year. I also acknowledge the help I received from my friends Andrew Chee, Jeffrey Hu and Justin Tzeng in completing my thesis.

WORKS CITED

Anand, Sudhir, and Paul Segal. "What Do We Know about Globalization?: Issues of Poverty and Income Distribution." Choice Reviews Online 45.09 (2008): n. pag. 2008. Web. 11 May 2016.

Bourguignon;, François, and Christian Morrisson. "Inequality among World Citizens: 1820-1992." The American Economic Review (n.d.): n. pag. Picketty PSE. Sept. 2002. Web. 11 May 2016.

Chotikapanich, Duangkamon, Rebecca Valenzuela, and Prasad D.S. Rao. "Global and Regional Inequality in the Distribution of Income: Estimation with Limited and Incomplete Data." Social Science Research Network, n.d. Web. 11 May 2016.

Dowrick, Steve, and Muhammad Akmal. "Contradictory Trends in Global Income Inequality: A Tale of Two Biases." Australian Bureau of Agriculture and Resource Economics. N.p., 2 June 2005. Web. 11 May 2016.

"GINI Index (World Bank Estimate)." Data. World Bank, n.d. Web. 11 May 2016.

Milanovic, Branko. "All the Ginis, 1950-2012." Research at the World Bank. World Bank, 2014. Web. 11 May 2016.

Milanovic, Branko. Global Income Inequality By the Numbers: In History and Now – An Overview -- (n.d.): n. pag. Heyman Center, 2012. Web. 11 May 2016.

"Remarks by the President on Economic Mobility." The White House. The White House, 04 Dec. 2013. Web. 11 May 2016.

Sala-i-Martin, Xavier. "The World Distribution of Income: Falling Poverty and … Convergence, Period." The Quarterly Journal of Economics (2006): n. pag. Columbia University, May 2006. Web. 11 May 2016.

"Society at a Glance 2014 | OECD READ Edition." OECD ILibrary. N.p., Dec. 2014. Web. 11 May 2016.

APPENDIX

Figure A: Estimates of Global Interpersonal Inequality at PPP$: Gini Coefficient, as compiled by Anand and Segal in *What Do We Know about Global Income Inequality?*

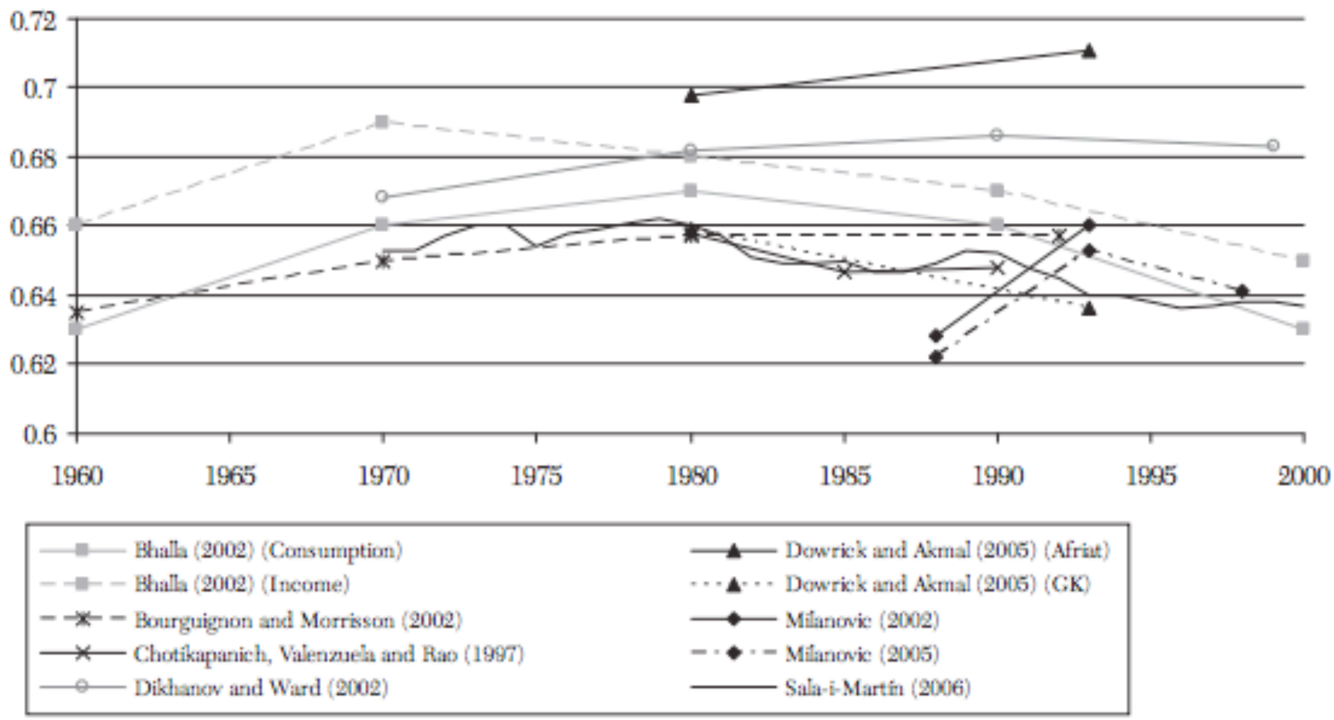

Figure 1. Estimates of Global Interpersonal Inequality at PPP$: Gini Coefficient

Figure B: Countries in the Model

Country Name	World Bank Country Code
Bangladesh	BGD
Brazil	BRA
Canada	CAN
China	CHN
Costa Rica	CRI
Germany	DEU
Denmark	DNK
Egypt, Arab Rep.	EGY
Spain	ESP

Ethiopia	ETH
Finland	FIN
France	FRA
United Kingdom	GBR
Georgia	GEO
Indonesia	IDN
India	IND
Iran, Islamic Rep.	IRN
Japan	JPN
Moldova	MDA
Mexico	MEX
Macedonia, FYR	MKD
Nigeria	NGA
Norway	NOR
Pakistan	PAK
Philippines	PHL
Poland	POL
Romania	ROU
Russian Federation	RUS
Tunisia	TUN
Turkey	TUR
Ukraine	UKR
United States	USA
Venezuela, RB	VEN
Vietnam	VNM

South Africa	ZAF

Figure C: 1990 Log X and Predicted Log X v/s Upper Bound of Income Bracket

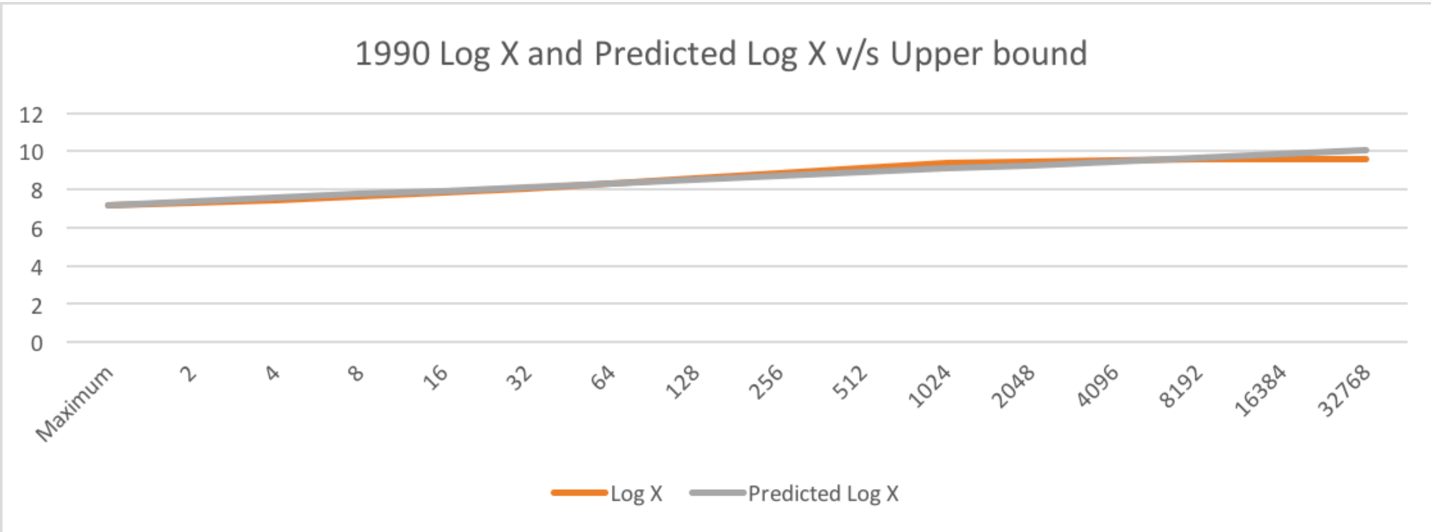

Figure D: 2000 Log X and Predicted Log X v/s Upper Bound of Income Bracket

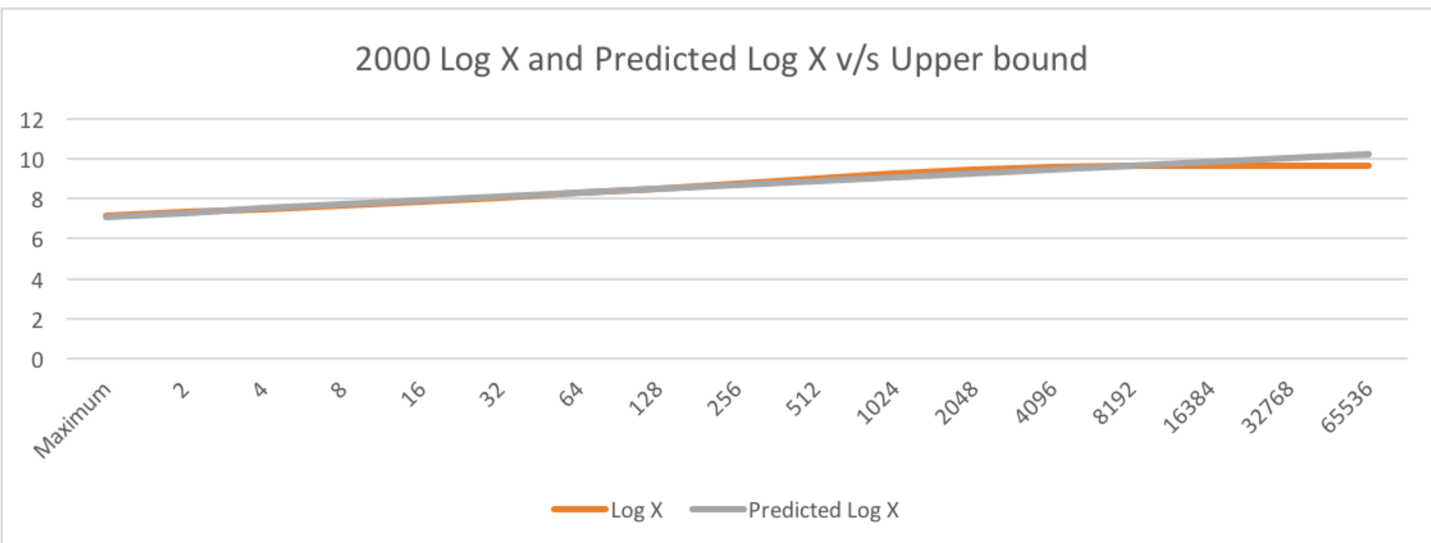

Figure E: 2010 Log X and Predicted Log X v/s Upper Bound of Income Bracket

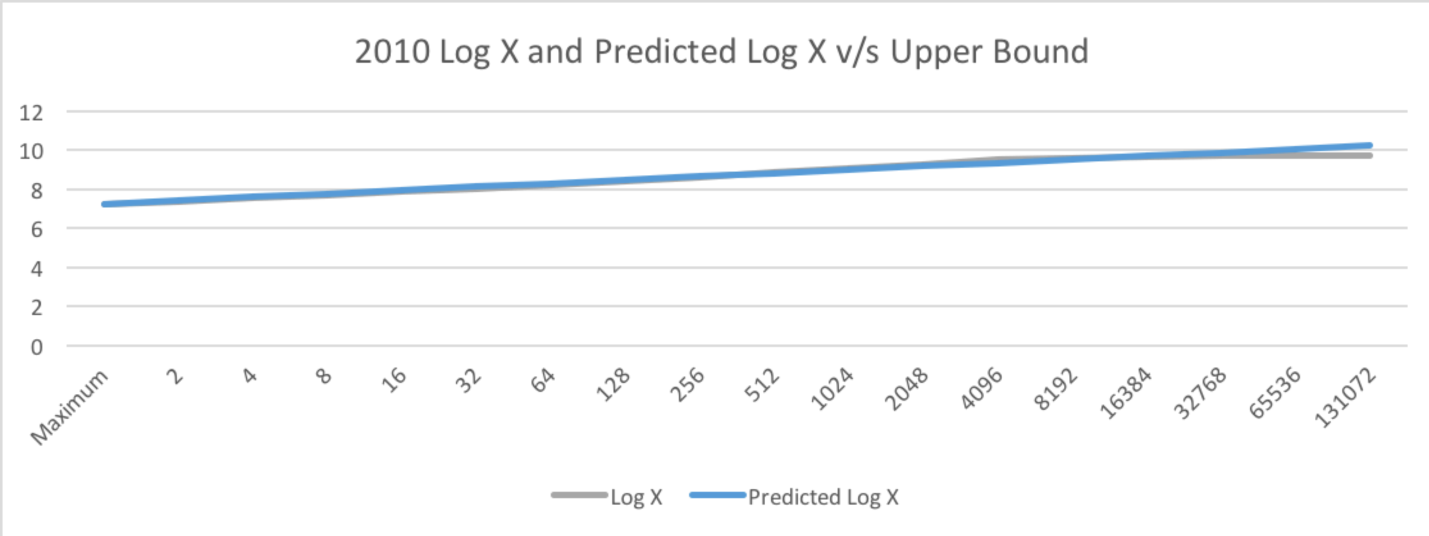

Figure F: Gini Coefficient over time

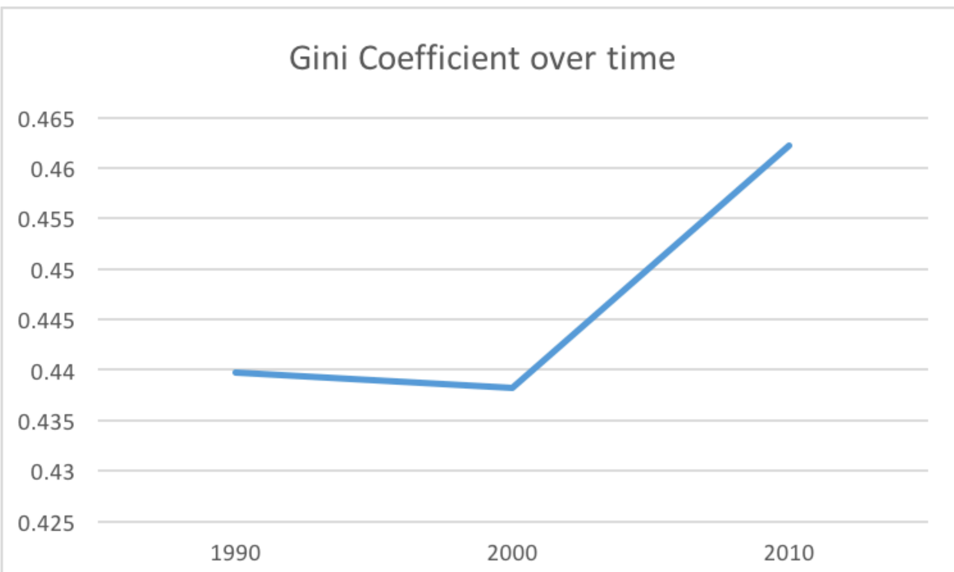

Figure G: Lorenz Curve over time

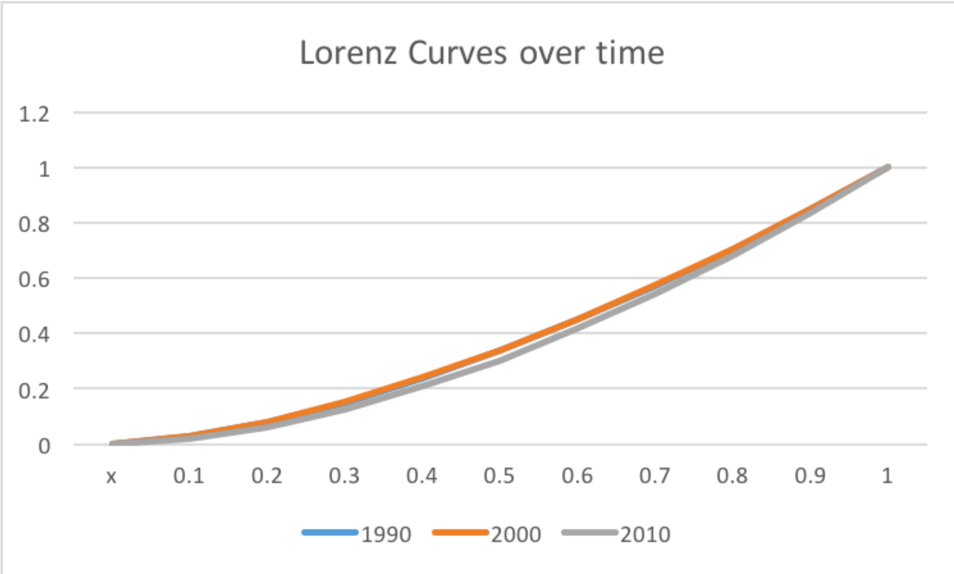

Figure H: Population weighted Gini coefficients by year

Year	Population-weighted Gini coefficients
1990	0.35
2000	0.38
2010	0.41