

A Model for Evaluating the Policy Impact on Poverty

Weishuang Qu and Gerald O. Barney

Millennium Institute

1117 North 19th Street, Suite 900

Arlington, VA 22209, USA

Phone/Fax: 703-841-0048/703-841-0050

Email: wqu@threshold21.com

Keywords: Poverty reduction, income distribution, Gini coefficient, T21 Model

I. Introduction

For decades poverty reduction in developing countries has been one of the most challenging tasks for the national policy maker. Recent trends show the income distribution in most countries has become more biased toward the rich. This paper reports on efforts to develop a dynamic model to evaluate the impact of alternative policies focused on the poor, including such common ideas as providing them free health care and reducing their income tax obligation. The model measures the impact of a particular policy on the Gini coefficient or on the number of households below an arbitrary poverty line. The purpose for creating the model is to possibly develop a poverty reduction sector to include in the T21 (Threshold 21) National Development Model.

The economic literature shows that household income distribution in most countries is close to the lognormal distribution. Hence the lognormal distribution was selected for this work to represent the income distribution, and a method for estimating the mean and standard deviation of the lognormal distribution is developed (see Section II). The method starts from household survey data of the country at a certain time. Over time, both the mean and the standard deviation change. It is straight-forward to model the change of the income mean using elements already in the T21 model, such as GDP, tax rate, household size, and population. To estimate the standard deviation, however, is more complicated. To make matters worse, it is difficult for most people to understand the meaning of the standard deviation, even when you have computed it, as, for example, \$15,000.

To provide an indicator that is easier for most people to understand than the standard deviation, we chose the Gini coefficient. In Section III, it is explained that the Gini coefficient is related to the ratio of the income standard deviation over the income mean. Then two methods are developed: one for deriving the income distribution from the known Gini coefficient and income mean, and the other for computing the Gini coefficient from a known lognormal distribution.

Section IV explains how the model takes government poverty reduction policy as input and generates changes to the Gini coefficient and the number of households under the poverty line as output, assuming all other factors, such as GDP, employment, and foreign aid, stay constant.

Section V demonstrates the model built in Vensim and Section VI lists a few topics for further research.

II. Developing the Lognormal Income Distribution from Household Survey Data

1. Household survey of income

Table 1 is a summary of a 1997-98 urban household income survey of a large developing country.

Income class	Lower bound	Upper bound	Number of households	Percentage
1	0	30000	10534	21.87%
2	30001	60000	16576	34.42%
3	60001	90000	10650	22.11%
4	90001	125000	5439	11.29%
5	125001	infinity	4962	10.30%

Table 1: Urban income survey summary

In the table there are five income classes (rows). The lower and upper income bounds of each class are specified in the second and third columns. The fourth column is the count of households in the class, and the fifth column is the percentage of households in each class.

2. Finding the lognormal distribution of income

We assume that household income is *lognormally* distributed, as is found to be the case in most countries. The lognormal distribution is defined by two parameters: mean and standard deviation. Thus, finding the lognormal distribution is equivalent to finding the mean and the standard deviation of the income.

The mean of income can be obtained from national accounts data, dividing national income by the number of households. If you do not have national accounts data for the country, you can still estimate the mean from Table 1 or from some other source.

Given the mean, a program we developed in C++ estimates the standard deviation of the lognormal income distribution. For the example in Table 1, the mean is estimated to be 65,033, and the standard deviation is 50,053.

Another C++ program was developed to calculate the percentage of households in each of the five income classes from the lognormal distribution with mean of 65,033 and standard deviation of 50,053. The results are listed in Column 6 of Table 2. Column 7 of Table 2 calculates the relative difference between data and model results. The first five columns of Table 2 are copied from Table 1.

Income class	Lower bound	Upper bound	Number of households	Percentage from data	Percentage calculated	Difference
1	0	30000	10534	21.87%	21.38%	-2.25%

2	30001	60000	16576	34.42%	37.46%	8.84%
3	60001	90000	10650	22.11%	20.50%	-7.30%
4	90001	125000	5439	11.29%	10.98%	-2.77%
5	125001	infinity	4962	10.30%	9.68%	-6.05%

Table 2: Comparing data with model results

The relative difference in Column 7 is calculated, using class 1 (first row) as example, as $(21.38\% - 21.87\%)/21.87\% = -2.25\%$. All differences are within the range of (-10%, 10%). Considering the error range in household income survey caused by many factors, such as bias in sampling and under- or over-reporting, we regard the differences in column 7 as acceptable.

3. Application of the lognormal income distribution

With the Lognormal Distribution, we have developed C++ programs to calculate several indicators, such as the fraction of income that the poorest fraction (from zero to 100%) of the population makes, and the fraction of households below any arbitrary poverty line (such as \$1/day per person). In the Table 1 example, the poorest 20% of households get 6.4% of income, and using income level of 40,000 as the household poverty line, 35.5% of households are below this line.

4. Limitation of this approach

Over time, both the mean and the standard deviation will change. It is straightforward to model and to explain the change of the income mean using elements already in the T21 model, such as GDP, tax rate, household size, and population. To estimate the standard deviation, however, is more complicated. To make matters worse, it is difficult to interpret the meaning of a standard deviation of, say 50,053, to an ordinary user, even when you have computed it. As explained in the next section, we chose the Gini coefficient as an income distribution indicator that most people can readily understand.

III. Linking Lognormal Distribution to the Gini coefficient

1. Relationship between Gini coefficient and the lognormal distribution

As we explained earlier, the lognormal distribution is defined by two parameters: *mean* and *standard deviation*. So when one thinks about the relationship between Gini and the lognormal distribution, one is in essence thinking about the relationship between Gini and the two parameters, the mean and the standard deviation.

It can be shown that, if income is lognormally distributed, there is a one-to-one correspondence between the Gini coefficient and the ratio of standard deviation over mean (S/M ratio).

For instance, if the income mean of a country is \$50,000, and its standard deviation is \$30,000, the Gini coefficient can be computed as 30.39%. The S/M ratio is computed as 30000/50000, or 0.6.

If we measure income in thousands of dollars, then the mean will be 50000/1000, or 50, and the standard deviation will be 30. But the S/M ratio will remain at 0.6. The Lorenz curve of the Gini coefficient will not change either, as the poorest $a\%$ ($0 < a < 100$) of population will have the same $b\%$ ($0 < b \leq a$) of income, no matter what unit you use to measure income.

We can also measure the income distribution in any other currency, Japanese Yen or Chinese RMB, by using an exchange rate. When converted to a different currency, both the mean and the standard deviation will be multiplied by the same parameter (i.e., the exchange rate), and their ratio will stay unchanged. The Lorenz curve will remain the same, and so will the Gini coefficient. We can even use an imaginary exchange rate to convert the income distribution to an imaginary currency, and the Gini coefficient will always remain the same.

Thus we can say, when the S/M ratio is 0.6, the Gini coefficient will be 30.39%. When the ratio changes, so does the Gini coefficient, and there exists a one-to-one correspondence between the two.

The one-to-one relationship is only true when income is log normally distributed. If the income distribution is non-lognormal, then one Gini coefficient value might correspond to multiple ratio values of standard deviation over mean.

2. Deriving income standard deviation from Gini coefficient and income mean

As there exists a one-to-one relationship between the S/M ratio and the Gini coefficient, a C++ program was developed to compute the table similar to the following to quantify their relationship.

S/M Ratio	0.1	0.2	0.5	1	1.5	2	3	4	5
Gini	5.18%	10.90%	26.05%	44.29%	55.54%	62.70%	71.03%	75.72%	78.74%

Table 3: Relationship between S/M ratio and Gini

The actual table has many more columns to define the relationship in more details.

Graphically, the table looks like Figure 1 below.

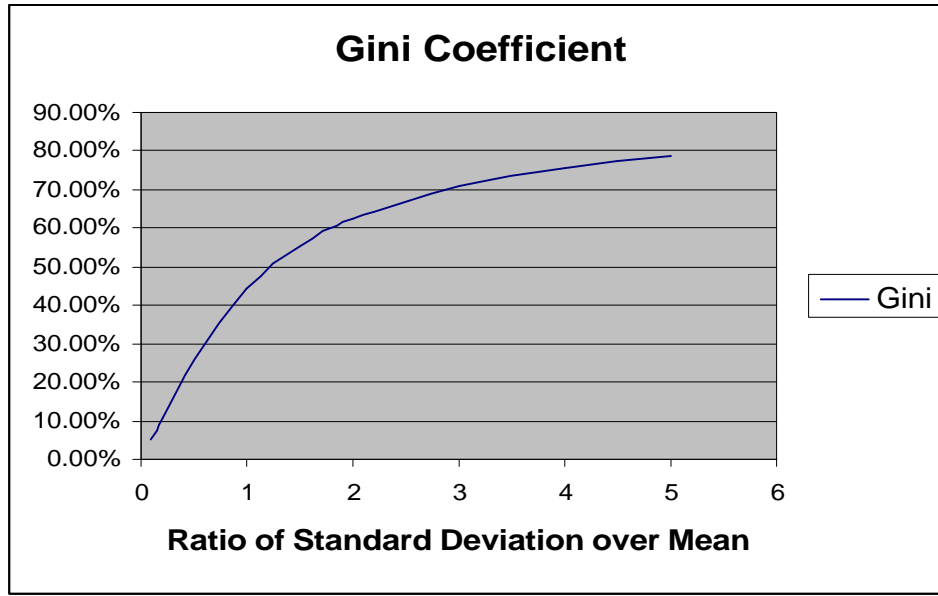


Figure 1: Graph of S/M Ratio and Gini Coefficient

With this table and using linear interpolation, we can quickly find the value of the standard deviation, given the values of the income mean and the Gini coefficient. For instance, if income mean is 3,000, and Gini coefficient is 30%, from the table (not Table 3, but the complete table with more details not presented here) and using linear interpolation, we can find that the S/M ratio is 59.09%. Then the value of standard deviation is:

$$S = M * S/M \text{ ratio} = 3000 * 59.09\% = 1772.7$$

3. Deriving Gini coefficient from a known lognormal income distribution

Similarly, with the table and using linear interpolation, we can find the Gini coefficient from the known values of the mean and the standard deviation of a lognormal income distribution.

We further found that the lognormal distribution is defined by any known point in the form of: “The poorest a% of population gets b% of income”. In other words, when that point is known, the S/M ratio is defined.

Since that point defines the ratio, it also defines the Gini coefficient. For many countries, there exists the estimate of the fraction of income that the poorest 20% of population gets. We can compute the Gini coefficient from that known point. A C++ program was developed and added to Vensim’s external function library to deal with the computation.

Although we can compute Gini coefficient from only one such point, it is advantageous to have multiple points, so that we can test whether the Gini coefficients computed from different points are similar. If they are not, it could indicate data error, or it could mean that the lognormal distribution is not a good fit to the income distribution of the country.

For the countries we have tested so far, the Gini coefficients derived from different points are all very close. Table 4 is China urban income data from the China Statistical Yearbook 2001 (Section 10-5).

Income % from low to high	10	10	20	20	20	10	10
Average disposable income	2653.02	3633.51	4623.54	5897.92	7487.37	9434.21	13311.02

Table 4: China urban income data

The table reads from left as: The lowest 10% of urban households has an average disposable income of 2653.02 RMB (per person per year), the next 10% has an average disposable income of 3633.51 RMB, and so on.

From Table 4 we calculated six points on the Lorenz curve, and further computed the six Gini coefficients from these points. The results are in Table 5. The Gini coefficients in the bottom row are surprisingly close.

Poorest fraction of Population	10.00%	20.00%	40.00%	60.00%	80.00%	90.00%
Income fraction	4.08%	9.66%	23.88%	42.01%	65.03%	79.54%
Gini coefficient computed	0.255	0.254	0.253	0.252	0.252	0.252

Table 5: Gini coefficients derived from China urban income data

The conclusions of this section are (1) that we can compute the Gini coefficient from a known lognormal distribution, (2) that we can even compute the Gini coefficient from a single point on the lognormal distribution, and (3) that with multiple points we can test data consistency and possibly adjust our estimates of the mean and the standard deviation of the lognormal income distribution.

4. Simulating income distribution dynamically under normal conditions

Most economists believe that the Gini coefficient of a country under normal conditions does not change much over a short to medium period of time. In other words, the S/M ratio remains rather constant over time. If we can obtain the value of Gini coefficient and assume that value stays constant, then when we simulate T21 into the future, we can easily compute the standard deviation from the endogenous mean of household income.

This method, of course, tells us nothing about the relative effectiveness of alternative policies to change the Gini coefficient.

In reality, however, Gini coefficient could, as a result of different policies, change noticeably even during a short period of time. Factors that might change the Gini coefficient include employment, government tax, government subsidy, and GDP growth. In the next section we will explore how changes in government policies can affect the Gini coefficient, assuming all other factors stay constant. Of course other factors will change, but in this first step, let's assume they do not.

IV. Policy Effect on Income Distribution

Let's assume that the government increases its income tax rate and uses the extra tax income to subsidize the poor in the form of providing free health care to the poor and free lunch to school children of families below the poverty line. If GDP is not affected, and if the government uses all the extra tax income on these subsidies, then the mean of household income should be the same. But the standard deviation, or the Gini coefficient, will be different.

Of course increasing the income tax rate may not be politically feasible, and GDP could be negatively affected by higher income tax rates. But as a first step, we would assume that these policies are doable, and GDP growth will not be affected.

Steps of the method are as follows:

1. Assume our interest is in the poorest 20% of population. Before implementing the intended government poverty reduction policies, we can calculate how much income this group of households makes based on the Lognormal Distribution. The result is, let's assume, 6.5% of total income. We can also compute the Gini coefficient based on the information that the lowest 20% population gets 6.5% of income. The result is 36.49%.
2. Calculate the real income this group of households will make, including health care and children's lunch provided free, after the implementation of the government policies. The result could be, let's assume, 6.7%. And the corresponding Gini coefficient would be 35.69%
3. As the mean of household income does not change, while the Gini coefficient does, we can compute the two standard deviation values from the same mean but different Gini values, corresponding to two different policy scenarios: adopting these policies or not.
4. With the mean and the standard deviation values of the log normal distributions known, we can compute and compare the poverty situation of the country before and after the policy implementation, such as the number of households under the poverty line.

V. Model in Vensim

An income and poverty reduction policy model built in Vensim will be demonstrated during presentation, and its run-only version will be available by request. Initial application of this model in the T21 national development model will be discussed.

VI. Further Research

Further study is needed in at least the following areas:

1. The feedback effect of government poverty reduction policies on GDP growth.
2. The feedback effect of poverty conditions on national development.
3. How other factors might change when government policy changes.

References:

1. Greene, William H. 1993. *Econometric Analysis*. 3rd Edition. (p.71) Prentice Hall.
2. Aitchison and Brown. 1969 *The Lognormal Distribution with Special Reference to its Use in Economics*. New York: Cambridge University Press,
3. Pearce, David W. 1994. *The MIT Dictionary of Modern Economics*. 4th Edition, (p. 254) The MIT Press.
4. Suhir, Ephrain. 1997. *Applied Probability for Engineers and Scientists*, (p.94). McGraw Hill.