A STUDY OF THE INCOME DISTRIBUTION UNDERLYING THE RASCHE, GAFFNEY, KOO AND OBST LORENZ CURVE

Duangkamon Chotikapanich

No. 63 - May, 1992
1 Introduction

One powerful method of illustrating the size distribution of income and wealth is the Lorenz curve. Two useful ways to construct the Lorenz curve are i) to first specify the distribution function of income and then derive a functional form for the Lorenz curve resulting from the specific distribution, ii) to represent the Lorenz curve directly by specifying a suitable function which satisfies all the properties of a well-behaved Lorenz curve.

For the first method of specifying the distribution function, there are two statistical distributions widely used as approximations to models of income distribution. They are the Pareto distribution and the lognormal distribution. Unfortunately, neither of them fits the entire range of the distribution well. For the second method of directly specifying a suitable function, many functional forms have been proposed in the literature. The performances of each functional form in terms of estimating the Gini coefficient, estimating the expenditure shares and estimating the Lorenz curve are evaluated extensively in Chotikapanich (1991). It was found that the functional form proposed by Rasche, Gaffney, Koo and Obst (1980) (RGKO) performed best in a number of ways. For this reason, if a distribution function can be derived from this particular Lorenz function it should provide a reasonably close approximation to the true distribution and hence a good model for the distribution of income.

This paper examines a possible distribution function that can be derived from the RGKO's Lorenz function. A suitable distribution function which fits the whole range of the distribution is developed.

2 The distribution function of income

The Lorenz curve which is the relationship between the cumulative proportion of income $x$, $\eta(x)$, and the cumulative proportion of income-receiving units, $\tau(x)$, is defined as

$$\eta = \eta(\tau)$$

$\tau(x)$, $\eta(x)$, and the relationship between $\eta$ and $\tau$ can be interpreted as follows. $\tau(x)$, in general, can be interpreted as a distribution function and can be
written as
\[ \pi(x) = \int_0^x f(r) \, dr \]  
(1)
where \( f(\cdot) \) is the density function of the income distribution. The function \( \pi(x) \) gives the cumulative proportion of population receiving income less than or equal to \( x \). The proportion of income earned by units whose incomes are less than or equal to \( x \), \( \eta(x) \), can be defined as
\[ \eta(x) = \frac{Q(x)}{\mu} \]  
(2)
where \( Q(x) \) and \( \mu \), the mean of the income distribution, are defined as
\[ Q(x) = \int_0^x r f(r) \, dr \]
\[ \mu = \int_0^\infty r f(r) \, dr \]
Using these definitions, we can write \( \eta(x) \) as
\[ \eta(x) = \frac{1}{\mu} \int_0^x r f(r) \, dr \]  
(3)
\( \eta(x) \) can then be interpreted as first moment distribution of \( x \).

One piece of information that is useful for explaining the distribution of income at different income level is the slope of the Lorenz curve. The slope of the Lorenz curve, which is \( \frac{d\eta}{dx} \), indicates how the proportion of income changes as proportion of population changes. By using \( \frac{d\eta}{dx} \) and \( \frac{d\pi}{dx} \) obtained from Equations (1) and (3) we have
\[
\frac{d\eta}{d\pi} = \frac{x}{\mu}
\]  
(4)
This piece of information will be used later in the analysis.

3 The distribution function for the RGKO’s Lorenz curve

In this section we will look at the possibility of obtaining the income distribution function derived from the RGKO’s Lorenz curve. It is found that the
full range of the distribution function can not be found, and so we consider the distribution functions suitable for different ranges of income and explore the possibility of deriving a combined distribution function for the whole range of income.

RGKO's function for the Lorenz curve is

\[ \eta = (1 - (1 - \pi)^\alpha) \frac{x}{\mu}, \quad \text{for } 0 < \alpha, \beta \leq 1 \] (5)

Differentiating Equation(5) by \( \pi \) we have

\[ \frac{d\eta}{d\pi} = \frac{\alpha}{\beta} (1 - \pi)^{\alpha - 1} (1 - (1 - \pi)^\alpha \frac{x}{\mu}) \frac{x}{\mu} \] (6)

Therefore, from Equations(4) and (6)

\[ \frac{x}{\mu} = \frac{\alpha}{\beta} (1 - \pi)^{\alpha - 1} (1 - (1 - \pi)^\alpha \frac{x}{\mu}) \frac{x}{\mu} \] (7)

This equation gives the relationship between \( \pi, x \) and \( \mu \), i.e., the distribution of income, the income and the mean income. It shows the distribution of income that is implied by RGKO's Lorenz curve at different income levels.

As Anstis (1978) has pointed out, the density function for the distribution defined in Equation(5) cannot in general be obtained explicitly in terms of standard functions. However, by using a Taylor series expansion on Equation(7), Anstis has derived a few approximations. Two of these approximations which will be useful in the later analysis within the context of this paper are as follows.

1. When \( \pi \) approaches one, and \( x > \frac{\alpha \mu}{\beta} \)

\[ \pi \approx 1 - \left( \frac{\beta x}{\alpha \mu} \right)^{\frac{1}{1 - \alpha}} \] (8)

and

\[ f(x) \approx \frac{1}{(1 - \alpha) \mu \alpha} \left( \frac{\beta x}{\alpha \mu} \right)^{-\frac{(2 - \alpha)}{1 - \alpha}} \] (9)
2. When \( \pi \) approaches zero and \( 0 < \alpha < 2\beta \) holds,

\[
\pi \approx 1 - \left(1 + \left(\frac{\beta x}{\alpha \mu}\right)^{\frac{\alpha}{\alpha - \beta}}\right)^{-\frac{1}{\beta}}
\]  
(10)

and

\[
f(x) \approx \frac{\frac{1}{\mu(1-\beta)} \left(\frac{\beta x}{\alpha}\right)^{\frac{2-\beta}{\beta}}}{\left(1 + \left(\frac{\beta x}{\alpha \mu}\right)^{1-\beta}\right)^{1+\frac{1}{\beta}}}
\]  
(11)

The distribution function(10) and the density function(11) are defined for values of \( x \) less than \( \frac{\alpha \mu}{\beta} \).1

Until now we could use this information to obtain the distribution function, \( \pi \), and the density function, \( f(x) \), only at the extreme ends of the distribution. However, it is worth investigating the possibility that the above two functions that explain the distribution at the upper and lower limits of the distribution adequately could be fused together (at some appropriate fuse point) so that a distribution that fits the whole range may be found. If so, then \( f(x) \) that denotes the density function over the entire range of \( x \) can be written as

\[
f(x) = f_1(x) , \quad \text{when} \quad x < x^* \]  
(12)

\[
= f_2(x) , \quad \text{when} \quad x \geq x^*
\]

where \( f_1(x) \) is the density function in Equation(11) and it can be written as

\[
f_1(x) = \frac{\frac{1}{\mu(1-\beta)} \left(\frac{\beta x}{\alpha}\right)^{\frac{2-\beta}{\beta}}}{\left[1 + \left(\frac{\beta x}{\alpha \mu}\right)^{1-\beta}\right]^{1+\frac{1}{\beta}}}
\]  
(13)

and \( f_2(x) \) is the density function in Equation(9) which can be written as

\[
f_2(x) = \frac{1}{(1-\alpha) \alpha \mu} \left(\frac{\beta x}{\alpha \mu}\right)^{\frac{-1}{1-\alpha}}
\]  
(14)

\footnote{Note that the valid range of values of \( x \) for the distribution function(10) and the density function(11) is slightly different from the one in Anstis (1978)}
$x^*$ is the cut-off point that has to be defined. Because $f_1(x)$ was originally derived as an approximation when $x < \frac{\alpha}{\beta}$ and $f_2(x)$ was originally derived as an approximation when $x > \frac{\alpha}{\beta}$, we write $x^* = \frac{\alpha c}{\beta}$. The constant $c$ is used to adjust the cut-off point to be somewhere around $\frac{\alpha c}{\beta}$ and it is unknown. The cut-off point will be at $\frac{\alpha c}{\beta}$ if $c = 1$. Note that $f(x)$ is discontinuous at $x = \frac{\alpha c}{\beta}$. However, it is common to use discontinuous functions when using different approximations at low and high income. See, for example, Ahmed and Bhattacharya (1974).

Two relevant and important properties of the density function (12) have to be explored. Firstly, if Equation (12) is to behave as a proper density function, it should have the property that

$$\int_0^\infty f(r)dr = 1$$

Because $f(x)$ in Equation (12) is made up of two functions, it is not necessary that the integral is equal to unity; in which case an adjustment is necessary to ensure that this property holds. Secondly, Equation (12) which is the density function of the income distribution must yield a mean equal to μ.

4 The properties of the combined density function

4.1 The integral of the density function

The crucial property that the density function must have is that the integral of the density function from zero to infinity must yield unity. To ensure that the combined density function (12) has this property, we define the density function $f^*(x) = \frac{1}{k}f(x)$ where $k$ is the normalization constant which ensures the condition that

$$\int_0^\infty f^*(x)dx = 1$$

(15)

From Equation (12)

$$\int_0^\infty f^*(x)dx = \int_0^{x^*} \frac{1}{k}f_1(x)dx + \int_{x^*}^\infty \frac{1}{k}f_2(x)dx$$

(16)

Further from Equation (13),
With simple evaluation of the definite integral, it can be shown that

\[ \int_0^\infty \frac{1}{k} f_1(x) \, dx = \frac{1}{k} (1 - (1 + c_{1-\alpha}^{-1})^{-\frac{1}{\alpha}}) \]

The second term on the right hand side of Equation (16) can similarly be written as

\[ \int_0^\infty \frac{1}{k} f_2(x) \, dx = \frac{1}{k} \frac{1}{\alpha \mu} \beta \left( \frac{\beta x}{\alpha \mu} \right)^{-\frac{1}{\alpha}} \]

Therefore from Equation (16), we have

\[ \int_0^\infty f^*(x) \, dx = \frac{1}{k} (1 - (1 + c_{1-\alpha}^{-1})^{-\frac{1}{\alpha}}) + \frac{1}{k} c_{1-\alpha}^{\frac{1}{\alpha}} \]

and the constant of integration is

\[ k = (1 - (1 + c_{1-\alpha}^{-1})^{-\frac{1}{\alpha}}) + c_{1-\alpha}^{\frac{1}{\alpha}} \] (17)

This result suggests that the combined density function in (12) has to be normalized using the value of \( k \) in Equation (17) to ensure that the area under the curve is unity.

### 4.2 The mean value

If \( f^*(x) \) is to have a mean that is consistent with the mean of the income distribution given in Section 2, the mean of the combined density function should also be \( \mu \). It is defined as
\[ \mu = E(x) = \int_0^\infty x f^*(-x) \, dx \]

From Equation (16), \( \mu \) can be written as

\[ \mu = E(x) = \int_0^{a_k} x f^*_1(x) \, dx + \int_{a_k}^\infty x f^*_2(x) \, dx \tag{18} \]

The first term on the right hand side of Equation (18) can be expanded as

\[
\int_0^{a_k} x f^*_1(x) \, dx = \int_0^{a_k} x \frac{1}{k} \left( \frac{1}{\mu (1-\beta)} \left( \frac{\beta_x}{\alpha} \right)^2 \left( \frac{\beta_x}{\alpha} \right)^{\frac{2\alpha - 1}{\alpha}} \right) \, dx \\
= \int_0^{a_k} \left( \frac{\beta_x}{\alpha} \right)^{\frac{1+\beta}{\alpha}} \left( 1 + \left( \frac{\beta_x}{\alpha} \right)^{\frac{1+\beta}{\alpha}} \right) \, dx \\
= \int_0^1 \frac{\mu}{k} \frac{1+\beta}{c^{1-\beta}} (1 + c^{1-\beta} z^\beta)^{-1 + \frac{1}{\alpha}} \, dz \\
= \frac{\mu}{k} \frac{1+\beta}{c^{1-\beta}} I 
\]

where \( z = \left( \frac{\beta_x}{\alpha mc} \right) \)

\[ I = \int_0^1 (1 + c^{1-\beta} z^\beta)^{-1 + \frac{1}{\alpha}} \, dz \]

\[ = \frac{\mu}{k} \frac{c^{1-\beta}}{c^{1-\beta}} b^* \frac{\Gamma(a^*) \Gamma(b^*)}{\Gamma(a^* + b^*)} (1 - I_d(a^*, b^*)) \tag{19} \]

where \( a^* = 1 + \frac{1}{\alpha} - \frac{1}{\beta} \)

\[ b^* = \frac{1}{\beta} \]

\[ d = (1 + c^{1-\beta})^{-1} \]
\(\Gamma(a^\ast), \Gamma(b^\ast)\) and \(\Gamma(a^\ast + b^\ast)\) are the gamma function for \(a^\ast\) and \(b^\ast\) respectively. \(I_d(a^\ast, b^\ast) = \frac{\Gamma(a^\ast + b^\ast)}{\Gamma(a^\ast)\Gamma(b^\ast)}B_d(a^\ast, b^\ast)\), where \(B_d(a^\ast, b^\ast)\) is the lower tail of the beta function of the first kind with parameters \(a^\ast\) and \(b^\ast\).

The second term on the right hand side of (18) can be expanded as

\[
\int_{x^\ast}^{\infty} x f_2(x) dx = \int_{x^\ast}^{\infty} \frac{x}{k} \left( \frac{1}{(1-\alpha)\alpha\mu} \left( \frac{\beta x}{\alpha\mu} \right)^{-\frac{(2-a)}{(1-\alpha)}} \right) dx
\]

\[
= \int_{x^\ast}^{\infty} \frac{1}{k} \left( \frac{\beta x}{\alpha\mu} \right)^{-\frac{1}{1-\alpha}} dx
\]

using transformation \(y = \frac{\beta x}{\alpha\mu}\), \(dx = \frac{\alpha\mu}{\beta} dy\) the integral is

\[
= \int_{c}^{\infty} \left( \frac{1}{k} \frac{1}{1-\alpha} y^{-\frac{1}{1-\alpha}} \right) dy \frac{\alpha\mu}{\beta}
\]

\[
= \int_{c}^{\infty} \frac{1}{k} \frac{1}{1-\alpha} \alpha\mu y^{-\frac{1}{1-\alpha}} dy
\]

\[
= \frac{1}{k} \frac{\mu}{\beta} c^{\frac{1-a}{\alpha}}
\] (20)

Substituting (19) and (20) into (18),

\[
\mu = E(x) = \mu E(x) = \frac{\mu}{k} c^{\frac{1-a}{\beta}} b^* \frac{\Gamma(a^\ast)\Gamma(b^\ast)}{\Gamma(a^\ast + b^\ast)} (1 - I_d(a^\ast, b^\ast)) + \frac{1}{k} \frac{\mu}{\beta} c^{\frac{1-a}{\alpha}}
\]

\[
= \mu \left[ c^{\frac{1-a}{\beta}} b^* \frac{\Gamma(a^\ast)\Gamma(b^\ast)}{\Gamma(a^\ast + b^\ast)} (1 - I_d(a^\ast, b^\ast)) + \frac{1}{k} \frac{\mu}{\beta} c^{\frac{1-a}{\alpha}} \right]
\]

\[
\mu = E(x) = \mu P
\] (21)

where \(P = \frac{1}{k} b^* \left[ c^{\frac{1-a}{\beta}} \frac{\Gamma(a^\ast)\Gamma(b^\ast)}{\Gamma(a^\ast + b^\ast)} (1 - I_d(a^\ast, b^\ast)) + \frac{1}{k} c^{\frac{1-a}{\alpha}} \right]\)

For Equation(21) to be the mean, \(P\) has to be equal to one. That is

\[
1 = \frac{1}{k} b^* \left[ c^{\frac{1-a}{\beta}} \frac{\Gamma(a^\ast)\Gamma(b^\ast)}{\Gamma(a^\ast + b^\ast)} (1 - I_d(a^\ast, b^\ast)) + \frac{1}{k} c^{\frac{1-a}{\alpha}} \right]
\]
or

\[ k = b^* \left[ c^{\frac{\alpha}{\beta}} \frac{\Gamma(a^*)\Gamma(b^*)}{\Gamma(a^* + b^*)}(1 - I_d(a^*, b^*)) + c^{\frac{1 - \alpha}{\beta}} \right] \]  

(22)

Equations (17) and (22) each express the unknown \( k \) as a function of \( c \). Therefore we can equate them to obtain the following equation in the unknown \( c \).

\[ (1 - (1 + c^{\frac{\alpha}{1 - \beta}})^{-\frac{1}{\alpha}}) + c^{\frac{1 - \alpha}{\beta}} = b^* \left[ c^{\frac{\alpha}{\beta}} \frac{\Gamma(a^*)\Gamma(b^*)}{\Gamma(a^* + b^*)}(1 - I_d(a^*, b^*)) + c^{\frac{1 - \alpha}{\beta}} \right] \]

or

\[ (1 - (1 + c^{\frac{\alpha}{1 - \beta}})^{-\frac{1}{\alpha}}) + c^{\frac{1 - \alpha}{\beta}} = c b^* \left[ c^{\frac{\alpha}{\beta}} \frac{\Gamma(a^*)\Gamma(b^*)}{\Gamma(a^* + b^*)}(1 - I_d(a^*, b^*)) + c^{\frac{1 - \alpha}{\beta}} \right] \]

and

\[ c = \frac{(1 - (1 + c^{\frac{\alpha}{1 - \beta}})^{-\frac{1}{\alpha}}) + c^{\frac{1 - \alpha}{\beta}}}{b^* \left[ c^{\frac{\alpha}{\beta}} \frac{\Gamma(a^*)\Gamma(b^*)}{\Gamma(a^* + b^*)}(1 - I_d(a^*, b^*)) + c^{\frac{1 - \alpha}{\beta}} \right]} \]  

(23)

Given the values for \( \alpha \) and \( \beta \), this equation can be solved in iterative fashion. We begin with initial value of \( c \), say \( c = 1 \), compute the right hand side for \( c = 1 \), and hence find a second value of \( c \). The right hand side is computed for the second value of \( c \) and the process is continued until convergence.

5 Graphical representation of the fitted distribution

Equation(12) involves the parameters \( \alpha \) and \( \beta \) and the unknowns \( k \) and \( c \). To be able to use it as a density function for any income distribution, these parameters must be estimated and the unknowns must be solved for. The parameters \( \alpha \) and \( \beta \) can be estimated using non-linear estimation of the Lorenz curve (5). For the unknown \( c \), the iterative procedure is performed on Equation(23). Once \( \alpha \) and \( \beta \) are estimated and \( c \) is solved for, \( k \) can be obtained from Equation(17). To obtain the cut-off point \( z^* = \frac{\alpha \mu}{\beta} \), estimation of \( \mu \) is the remaining problem. The sample mean \( \bar{z} \) is used as the estimator.
In this section, a set of data on the 1981 expenditure of 572 households in municipal areas in the Northern part of Thailand is used to obtain and to plot the combined density function $f^*(x)$ and the corresponding combined distribution function. This set of data is the total data for that area obtained from the 1981 nationwide Socio-economic Survey conducted by the National Statistical Office of Thailand.

By using non-linear estimation of RGKO Lorenz curve (5) the $\alpha$ and $\beta$ are estimated as 0.55031 and 0.78611 respectively. The unknown $c$ is solved for by using the starting value of $c = 1$ in Equation(23). The iterative procedure converges at $c = 1.03$. By using $c = 1.03$, $k$ is calculated as 1.6799. The estimate for $\mu$ was $\bar{x} = 1193.66$. Given these above values of $\alpha$, $\beta$, $c$, $\mu$ and $k$ the combined density function $f^*(x)$ and the corresponding distribution function can be plotted.

For the convenience of the plotting, the combined function is plotted using the value of $f^*(x)$ against $\frac{\bar{z}}{\bar{\alpha} \cdot \mu}$ instead of $z$. To plot using $x$ values, the observed range from the actual data is very broad, from 192.19 to 10,984.86, while the range for $y = \frac{\bar{z}}{\bar{\alpha} \cdot \mu}$ is between 0.23 to 13.26. The smaller scale on $y$ is easier to handle in terms of plotting. In addition, since $y$ is a linear transformation of $x$, the distribution of $y$ is essentially the same as that of $x$.

The graphical results plotted using this adjusted combined functions are shown in Figures 1 to 3. The $\bullet$ is used in plotting the observed frequencies and $\ast$ is used in plotting the expected frequencies based on the combined functions. Figures 1, 2 and 3 show the plots of the distribution function, the density function and the density function at the lower range of expenditure distribution. It can be seen from these three figures that the entire adjusted combined functions are very close to the true functions. Thus, the adjusted combined functions should represent the income distributions very well. Figure 2 shows that the entire density function exhibits a bimodal property. This is particularly evident from Figure 3 that shows more details over the range 0.2 to 1.0.
Figure 1: Combined distribution function based on $f^*(x)$
Figure 2: Combined density function based on $f^*(x)$
Figure 3: Combined density function
(Enlarged portion over the range 0.2 and 1.0)
6 Conclusion

This paper has investigated the nature of the distribution and density functions that are consistent with the RGKO's Lorenz curve. It is found that the density function and the distribution function for the RGKO's equation cannot in general be obtained explicitly in terms of standard functions. However, it is demonstrated that a reasonable approximation of the distribution can be obtained which is consistent with Lorenz curve defined by RGKO. In view of the emphasis placed on the RGKO's equation in empirical studies, an examination such as the one undertaken in this paper is essential in understanding the characteristics of the underlying distribution. While the techniques employed in the derivation could be improved further, this paper demonstrates the feasibility of this kind of study.

7 References


A Note on A Bayesian Estimator in an Autocorrelated Error Model. William Griffiths and Dan Dao, No. 3 - April 1979.


Bayesian Econometrics and How to Get Rid of Those Wrong Signs. William E. Griffiths, No. 31 - November 1987.


A Generalized Theil-Tornqvist Index for Multilateral Comparisons.


A Comparison of Alternative Functional Forms for the Lorenz Curve.
Duangkamon Chotikapanich, No. 60 - October 1991.

A Disequilibrium Model of the Australian Manufacturing Sector.