

# Derivation of Theil's Inequality Measure from Lorenz Curves

Nicholas Rohde  
School of Economics  
The University of Queensland  
Brisbane, Australia

March, 2007

## *Abstract*

Theil's T inequality measure is a commonly used tool for inequality measurement as it satisfies important axiomatic requirements such as the principle of transfers and decomposability. Derived from the concept of information theory, Theil's measure represents a special case of the Generalised Entropy (GE) type measure. The main objective of this paper is to establish a relationship between Theil's inequality measure and the Lorenz curve, thus providing a useful link between these two popular tools for studying inequality. The result established in this paper also provides a basis for studying generalised entropy measures for inequality measurement using Lorenz curves. Analytical expressions for Theil's information theoretic measures are derived for three popular specifications of Lorenz curves. Empirical estimates of Theil's measures for these Lorenz curves are also presented for a number of cases.

JEL Classification: C10, D63, I32

Key words: Theil's T statistic, Lorenz curves, inequality measurement

Acknowledgements: I would like to thank Prasada Rao for his assistance in writing this article.

## 1. Introduction

The Lorenz curve is perhaps the most fundamental tool used to measure income inequality. Graphically, the Lorenz curve gives the proportion of total societal income accruing to the lowest earning proportion of income earners. This representation forms the basis of many common inequality measures. The Gini coefficient, for instance, is defined as twice the area between the considered Lorenz curve and the line of perfect equality, while Kakwani (1980) proposed a measure derived from the length of the Lorenz curve. Other inequality measures such as the Schultz coefficient or Robin Hood index are also derived directly from the Lorenz curve. As these measures are defined in terms of the Lorenz curve, they are frequently expressed directly in terms of the parameters dictating the shape of the curve.

Unlike the Gini coefficient and other inequality measures above, Theil's (1967) T statistic does not rely on the Lorenz curve for estimation. Instead it has its origin in information theory and is usually calculated in discrete form from tabulated income share data, or may be determined from the density function of the underlying income distribution. As a consequence, Theil's measure is not reported alongside other inequality metrics when parameter estimates for the Lorenz curve are given.

Theil's measure is part of a special class of inequality measures known as Generalised Entropy, or GE measures. Derived from the concept of information theory, Theil's measure seeks to quantify the level of disorder within a distribution of income. Cowell (1995) shows that any inequality measure that satisfies a particular axiomatic framework may be classed as a GE measure. An important property of Theil's measure and the class of Generalised Entropy inequality measures is the additive decomposability characteristic, which implies that the aggregate inequality measure can be decomposed into inequality *within* and *between* any arbitrarily defined population subgroups. Other common GE measures besides Theil's measure include the mean logarithmic deviation, half the square of the coefficient of variation and Atkinson's (1970) measure of inequality.

Alongside the Gini coefficient, Theil's measure remains the most popular of the class of GE measures and is employed in several influential studies on global income distributions including the recent works of Milanovic (2002), Sala-i-Martin (2003) and Chotikapanich, Rao and Tang (2006).

The main objective of this paper is to provide a formal and direct link between Theil's measure and the Lorenz curve. This is achieved by showing that Theil's T measure can be expressed in terms of the derivatives of the Lorenz curve. Once an algebraic relationship is established, it is possible to determine a general rule for calculating Theil's inequality measure from a parametrically specified Lorenz curve instead of deriving it from grouped data or the density function of the underlying distribution. The paper also provides analytical expressions for the Theil's measure associated with three commonly used specifications for the Lorenz curves. Parametric specifications are provided for the Kakwani-Podder (1973), Gupta (1984) and Chotikapanich (1993) Lorenz curves.

This paper is divided into five major sections. Section 2 provides a quick overview of the notation and techniques used within the paper. Section 3 discusses the discrete version of Theil's measure and shows that the discrete version and the more general formula in its continuous form can be expressed as functions of the Lorenz curve. Section 4 applies this formula to three simple and commonly used Lorenz curves and provides analytical expressions for the Theil's measure in terms of the parameters of the underlying Lorenz curves. Section 5 provides results from a simulation study to illustrate the link empirically.

## *2. Notation and basic concepts*

The paper considers the distribution of income as it accrues unevenly across a population of  $j$  individuals. We assume that all incomes are non-negative, and denote the income of the  $k$ th individual to be  $x_k$ . If the population is arbitrarily partitioned such that we have  $n$  groups, the income share and population share of the  $i$ th group are denoted  $p_i$  and  $q_i$  respectively. The income share of group  $i$  may be calculated as the total income accruing

to persons within income group  $i$ , divided by the total income of the population. Similarly the population share of group  $i$  is the proportion of total population contained within that group. For income group  $i$  that contains  $b$  individuals, the income and population shares may be calculated as

$$p_i = \frac{\sum_{k(i)=1}^b x_{k(i)}}{\sum_{k=1}^j x_k} \qquad q_i = \frac{b}{j}$$

Clearly  $\sum_{i=1}^n p_i = 1$   $\sum_{i=1}^n q_i = 1$

If income data is given in the form of population shares and corresponding income shares, Theil's T measure is easily calculated using the formula

$$(1) \qquad T = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i}$$

The statistic uses the expected information content of the income distribution to measure the level of inequality. Theil's measure may also be expressed as

$$(2) \qquad T = \ln(n) - S$$

where  $S$  is the Shannon entropy or information content of the distribution, which has a range from 0 to  $\ln(n)$ . Shannon entropy may be calculated for individual incomes or  $n$  income shares. For  $j$  individual incomes the information content may be calculated as

$$(3) \qquad S = \sum_{k=1}^j p_k \ln \frac{1}{p_k}$$

where  $p_k$  is the income share of the  $k$ th individual.

If all incomes are equalised, the information content in the distribution of income is maximised, causing  $S$  to take on the value of  $\ln(j)$ . Similarly if all but one  $p_k = 0$  and the remaining  $p_l = 1$ , the information content of the distribution is minimised. This allows Theil's measure to fall between 0 in the case of perfect equality and  $\ln(j)$  for perfect inequality.

Theil's measure may also be calculated continuously from the income density function  $f(x)$ . This represents the limiting case when the number of individual incomes  $j$ , or income groups  $n$  is extended infinitely. This is calculated as

$$(4) \quad T = \int_0^{\infty} \frac{x}{\mu} \ln\left(\frac{x}{\mu}\right) f(x) dx$$

where  $\mu$  is the mean income level. It is important to note that the continuous form of the Theil statistic lies on the interval  $[0, \infty)$ . In this form the Theil measure's information content interpretation is diminished as the concept does not apply to real numbers.

If the individual incomes are placed in ascending order such that  $x_1 < x_2 \dots < x_j$  a Lorenz curve may be constructed. The Lorenz curve is a plot of the cumulative income share of the lowest earning  $k$  individuals against the cumulative population share of the same group. If expressed as a continuous function, the Lorenz curve is typically given as

$$(5) \quad \eta = f(\pi)$$

where  $\pi$  is the cumulative population share and  $0 \leq \pi \leq 1$   
 $\eta$  is the cumulative income share and  $0 \leq \eta \leq 1$

and  $\frac{d\eta}{d\pi} > 0$ ,  $\frac{d^2\eta}{d\pi^2} > 0$ ,  $\eta(0) = 0$ ,  $\eta(1) = 1$

The Lorenz curve remains one of the most intuitive techniques applied to the measurement of inequality and forms the basis for a number of popular inequality measures.

### 3. Income shares and the Lorenz curve

We start with a linear piecewise Lorenz curve constructed from  $n$  equal sized segments, each describing successive income blocks. Data is often presented in this form, usually with  $n = 10$  equal sized income groups. If decile shares are to be used, they are represented as  $(\eta_1, \eta_2, \dots, \eta_{10})$  and are plotted against population shares  $(\pi_1, \pi_2, \dots, \pi_{10})$  to form the Lorenz curve. Clearly the point  $(\pi_{10}, \eta_{10})$  represents the termination point (1,1) on the Lorenz curve, while the origin would be represented as  $(\pi_0, \eta_0)$ . The population share of income group  $i$ , denoted  $q_i$ , is given as  $q_i = \pi_i - \pi_{i-1}$  where  $i$  is the income group under consideration. If we are dealing with decile data, each  $q_i = 0.1$ . The income data may be similarly disaggregated using the formula  $p_i = \eta_i - \eta_{i-1}$  to give income shares corresponding to each population share. Theil's T measure may then be calculated from the income and population shares using the formula

$$(6) \quad T = \sum_{i=1}^n p_i \ln \left( \frac{p_i}{q_i} \right)$$

where

$p_i$  is the income share of group  $i$

$q_i$  is the population share of group  $i$

The task now is to express Theil's measure such that it can be interpreted as a discrete form integral or Riemann sum. The Riemann sum approximates a definite integral by measuring the area under a curve. This is done by filling the space under the curve with a series of rectangles and summing these to measure the enclosed area. These rectangles may be infinitesimal in size, yielding a good approximation to the area under a curve. If

the area under the curve ranges from zero to one (as for a Lorenz curve) and is divided into  $n$  evenly sized partitions, the Riemann sum may be written as

$$(7) \quad \int_0^1 f(x)dx \approx \sum_{i=1}^n f(x_i) \frac{1}{n}$$

where  $f(x_i)$  gives the height of partition  $i$  and  $\frac{1}{n}$  gives the width. The product of these two terms gives the area of the enclosed rectangle; the sum of these terms gives an approximation of the total area under the curve. Equation (6) may be expressed as an approximate form integral by multiplying and dividing through by the scaling factor  $\frac{1}{n}$  to give

$$(8) \quad T = \sum_{i=1}^n \frac{1}{n} \left( \frac{p_i}{\frac{1}{n}} \right) \ln \left( \frac{p_i}{q_i} \right)$$

It should be evident that with  $n$  evenly sized (in terms of population) income groups under consideration, the population share of each group,  $q_i$ , is equal to  $\frac{1}{n}$ . Replacing this term in equation (8) gives the formula

$$(9) \quad T = \sum_{i=1}^n \frac{1}{n} \left( \frac{p_i}{q_i} \right) \ln \left( \frac{p_i}{q_i} \right)$$

Equation (9) is now of the form given for the Riemann sum integral in equation (7). Here

$\left( \frac{p_i}{q_i} \right) \ln \left( \frac{p_i}{q_i} \right)$  may be interpreted as the height of each column, given as  $f(x)$  in equation

(7), and  $\frac{1}{n}$  represents the width. Summing these areas gives an approximation of the

integral of the function  $\left( \frac{p_i}{q_i} \right) \ln \left( \frac{p_i}{q_i} \right)$ . The lower bound for this approximation is zero;

the upper bound at  $\frac{n}{n} = 1$ . An attractive property of the Riemann sum is that the limit of the sum as  $n \rightarrow \infty$  is the definite integral. This allows us to dispense with the approximately equals symbol in equation (7) when considering the limit, and allows us to refer to the enclosed area as a Riemann integral.

Our focus now turns to the limit of the ratios in equation (9) as  $n \rightarrow \infty$ , or the behaviour of Theil's T measure as the number of income groups increases.

As we obtained the income and population shares  $p_i$  and  $q_i$  from disaggregating the Lorenz curve, these shares may be interpreted as the "rise" and "run" of each piecewise segment. As we increase the number of segments to the Lorenz curve we get a parallel to Newton's Difference Quotient for the derivative.

This is

$$(10) \quad f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

which states that the derivative of  $f$  at  $x$  is given by the limit of the difference quotient as  $h \rightarrow \infty$ . This formula may be applied to the Lorenz curve. If two points are chosen on the  $\pi$  axis, the "run" or gap between the two points is given as  $\pi_i - \pi_{i-1}$ , which is denoted  $h$  in the difference quotient formula, and defined as the population share  $q_i$ . Similarly the "rise" is expressed as  $f(x+h) - f(x)$  which is given by  $\eta_i - \eta_{i-1}$  in Lorenz curve notation and is equal to income share  $p_i$ . Substituting these expressions into equation (10) produces the result that the derivative of  $\eta$  at  $\pi$  is

$$(11) \quad \eta'(\pi_i) = \lim_{q \rightarrow 0} \frac{p_i}{q_i}$$

This result is consistent with Kakwani's (1980) finding that the slope of the Lorenz curve is equal to  $\frac{x(F)}{\mu}$ , where  $x(F)$  is the inverse function of the income CDF and lies on the open interval  $[0,1)$ . Take the discrete income data outlined in section 2, where we have  $j$  non-negative incomes ordered such that  $x_1 < x_2 \dots < x_j$  where  $x_k$  represents the income accruing to the  $k$ th individual.

If  $F$  is taken to be equal to  $\frac{k}{j}$ , then  $x(F)$ , which may be interpreted as the income accruing to the person earning at the  $\frac{k}{j}$  proportion of income earners will be equal to  $x_k$ .

The mean income level is  $\mu = \frac{1}{j} \sum_{k=1}^j x_k$ , giving the result that the slope of the Lorenz curve using discrete data at point  $\pi = \frac{k}{j}$  is equal to

$$(12) \quad \eta' = \frac{x_k}{\left( \frac{1}{j} \sum_{k=1}^j x_k \right)}$$

The result used in this paper shows that the derivative of a linear segment is equal to  $\frac{p_i}{q_i}$ .

$p_i$  may be calculated from discrete data using the formula

$$(13) \quad p_k = \frac{x_k}{\sum_{k=1}^j x_k}$$

where the numerator is the income earned by income "group"  $x_k$  and the denominator is the total population income.

The population share is given as

$$(14) \quad q_k = \frac{1}{j}$$

Combining equations (13) and (14) gives the result that the slope of the Lorenz curve at  $\pi = \frac{k}{j}$  is equal to

$$(15) \quad \eta' = \frac{jx_k}{\left(\sum_{k=1}^j x_k\right)}$$

which is clearly equal to the result given by Kakwani in equation (12).

Applying results from equations (7) and (11) to equation (9) gives the result

$$(16) \quad T = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{n} \left(\frac{p_i}{q_i}\right) \ln\left(\frac{p_i}{q_i}\right) = \int_0^1 \eta' \ln \eta' d\pi$$

which is the general formula for the Theil T measure in terms of the Lorenz curve.

We now seek to express the right hand side of equation (16) in a more useful format.

The continuous Theil T statistic may be given as

$$(17) \quad T = \int_0^1 \eta' \ln(\eta') d\pi$$

Integrating (17) by parts gives

$$(18) \quad T = \eta \ln(\eta') - \int_0^1 \eta (\ln(\eta'))' d\pi$$

A property of the derivatives of logarithmic functions is that  $\frac{d}{dx} \ln(g(x)) = \frac{g'(x)}{g(x)}$ , thus the final term on the right of equation (18) can be written as  $\eta \left( \frac{\eta''}{\eta'} \right)$ . This gives right the equation for the Theil statistic in terms of the Lorenz curve as

$$(19) \quad T = \eta \ln(\eta') \Big|_0^1 - \int_0^1 \frac{\eta \eta''}{\eta'} d\pi$$

By substituting in the bounds of integration this may be simplified to

$$(20) \quad T = \ln(\eta') \Big|_{\eta=1} - \int_0^1 \frac{\eta \eta''}{\eta'} d\pi$$

Unfortunately further simplification is not practical, which leaves equations 17 and 20 as the general formulas for Theil's measure in terms of the Lorenz curve.

#### 4. Calculating analytic expressions for Theil's measure for simple Lorenz curves

In this section we derive analytical expressions for the Theil inequality measure associated with three selected functional specifications for Lorenz curves commonly used in the literature. The functional specifications are taken from Chotikapanich (1993), Gupta (1984) and Kakwani and Podder (1973). Equations (17) and (20) are used, which provide expressions of Theil's inequality measure in terms of the Lorenz curve and its first and second order derivatives.

##### *The Chotikapanich (1993) specification*

The Chotikapanich (1993) Lorenz curve is a simple functional form that has some useful properties. This functional form is particularly attractive for this purpose as it is easily integrable.

The functional form for the specification of the Lorenz curve,  $\eta = L(\pi)$  is

$$(21) \quad \eta = \frac{e^{k\pi} - 1}{e^k - 1} \quad k > 0$$

The first and second order derivatives of the Lorenz curve with respect to  $\pi$  are given by:

$$(22) \quad \eta' = \frac{ke^{k\pi}}{e^k - 1} > 0 \quad \eta'' = \frac{k^2 e^{k\pi}}{e^k - 1} > 0 \quad k > 0$$

From (22), the derivative of  $\eta$  evaluated at  $\pi = 1$  is given by  $\frac{ke^k}{e^k - 1}$

The first term in (20) is therefore equal to

$$(23) \quad \ln(\eta') \Big|_{\eta=1} = k + \ln k - \ln(e^k - 1)$$

The second term is equal to

$$(24) \quad \int_0^1 \frac{(e^{k\pi} - 1)(e^k - 1)(k^2 e^{k\pi})}{(e^k - 1)(ke^{k\pi})(e^k - 1)}$$

which can be simplified to

$$(25) \quad \int_0^1 \frac{k(e^{k\pi} - 1)}{(e^k - 1)}$$

which equals

$$(26) \quad \frac{e^{k\pi} - k\pi}{e^k - 1} \Big|_{\pi=0}^{\pi=1} \\ = \frac{e^k - k}{e^k - 1} - \frac{1}{e^k - 1}$$

Thus the Theil index for the Chotikapanich (1993) form of Lorenz curve is given by

$$(27) \quad T = k + \ln k - \ln(e^k - 1) - \left( \frac{e^k - k - 1}{e^k - 1} \right)$$

*The Gupta (1984) Lorenz curve*

Gupta (1984) provides another single parameter Lorenz curve given by the following equation.

$$(28) \quad \eta = \pi a^{\pi-1} \quad a > 0$$

It can be seen that the first and second order derivatives of the Lorenz curve in (28) are given by

$$\eta' = a^{\pi-1}(1 + \pi \ln a) > 0 \quad \eta'' = a^{\pi-1} \ln a(2 + \pi \ln a) > 0 \quad a > 0$$

Substituting into equation (20) we have

$$(29) \quad T = \ln(1 + \ln a) - \int_0^1 \frac{\pi a^{\pi-1} * a^{\pi-1} \ln a(2 + \pi \ln a)}{a^{\pi-1}(1 + \pi \ln a)} d\pi$$

The above definite integral is not able to be calculated analytically. The computer software package Mathematica is able to produce a series approximation of the result. It is given as

$$(30) \quad \int \frac{\pi a^{\pi-1} * a^{\pi-1} \ln a(2 + \pi \ln a)}{a^{\pi-1}(1 + \pi \ln a)} d\pi = \left( \frac{a^{x-1}x}{\ln a} - \frac{E_i(x \ln a + 1)}{ae \ln^2 a} \right) \ln a$$

where  $E_i$  is an exponential integral. An exponential integral is defined as

$$(31) \quad E_i(x) = - \int_{-x}^{\infty} \frac{e^{-t}}{t} dt$$

No analytic solution for the exponential integral exists, however it is useful for generating approximates to analytical solutions for integrals as many functions that do not possess an anti-derivative may be expressed in its terms. It may also be represented as the below series

$$(32) \quad E_i(x) = \gamma + \ln x + \sum_{k=1}^{\infty} \frac{x^k}{kk!}$$

where  $\gamma$  is the Euler gamma constant and is approximately equal to 0.57721. This constant is defined as the limiting difference between the harmonic series and the log function.

Combining (31) and (32) we can provide an expression for the Theil statistic associated with the Lorenz curve proposed by Gupta (1984). This is

$$(33) \quad T = \ln(1 + \ln a) - \left[ \left( \frac{1}{\ln a} - \frac{\gamma + \ln(\ln a + 1) + \sum_{k=1}^{\infty} \frac{(\ln a + 1)^k}{kk!}}{ae \ln^2 a} \right) \ln a + \left( \frac{\gamma + \sum_{k=1}^{\infty} \frac{1}{kk!}}{ae \ln^2 a} \right) \ln a \right]$$

where  $a$  is the parameter of the Lorenz curve and  $\gamma$  is the Euler gamma constant. Once the value of  $a$  is known, it is possible to compute the numerical value of T.

#### *Kakwani and Podder (1973) Lorenz curve specification*

Kakwani and Podder (1973) proposed the following functional form for the Lorenz curve which was utilised in their analysis of the income distribution data from the 1973 Australian income study.

The Kakwani-Podder (1973) Lorenz curve is specified as

$$(34) \quad \eta = \pi * e^{-\beta(1-\pi)} \quad \beta > 0$$

It is easy to show that the first and second order derivatives of the Lorenz curve are given by

$$(35) \quad \eta' = e^{-\beta(1-\pi)}(1 + \pi\beta) > 0 \quad \eta'' = \beta e^{-\beta(1-\pi)}(2 + \pi\beta) > 0 \quad \beta > 0$$

Evaluating the first term on the right of equation (34) at  $\pi = 1$  gives

$$(36) \quad \ln(\eta') = \ln(1 + \beta)$$

The integral of the second term is

$$(37) \quad \int \frac{\pi\beta(2 + \pi\beta)}{(1 + \pi\beta)} e^{-\beta(1-\pi)} = \frac{e^{-\beta-1}(\beta e^{\beta\pi+1}\pi - E_i(\beta\pi + 1))}{\beta}$$

Substituting expressions (35) and (36) in equation (20), the expression for the Theil T statistic for the Kakwani-Podder (1973) Lorenz curve may be given as

$$(38) \quad T = \ln(1 + \beta) - \left[ \frac{e^{-\beta-1} \left( \beta e^{\beta+1} - \left[ \left( \gamma + \ln(\beta + 1) + \sum_{k=1}^{\infty} \frac{(\beta + 1)^k}{kk!} \right) - \left( \gamma + \sum_{k=1}^{\infty} \frac{1}{kk!} \right) \right] \right)}{\beta} \right]$$

with  $\beta > 0$  as the parameter of the Lorenz curve. For purposes of computing the Theil index, it is necessary to approximate the infinite series on the right side of equation (32).

##### *5. A simulation based evaluation of Theil indexes computed in their Lorenz curve formulation*

In this section we examine the empirical validity of the expressions derived in Sections 3 and 4 using a simulation experiment. As the expressions for the Theil statistic provided in Section 4 are based on the relationship derived in Section 3, the expressions provided in equations (33) and (38) also require approximations for the series involved.

The simulation experiment used here is as follows.

1. For each of the Lorenz curve specifications due to Chotikapanich, Gupta and Kakwani-Podder, income shares of decile groups are generated for different parameter values.
2. Using the income shares for the decile groups, Theil statistics are computed using the discrete version of Theil's measure applied to discrete data. These values represent what is commonly reported as Theil's T statistics
3. For each of the parameter values for the Lorenz curves, values of the Theil statistic are computed using the analytical expressions derived in Section 4. The values reported here are based on the formal relationship between Theil's measure and the Lorenz curves established in this paper.

The validity and accuracy of the expressions derived in this paper will be reflected in the closeness of the numerical values of Theil's statistic computed using the simulated data as described in the three steps outlined above.

The following table shows the results from the simulation for six different parameter values for the Chotikapanich, Gupta and Kakwani and Podder Lorenz curves.

Column (2) of the table shows the estimates of the Theil statistic based on income share data for the decile groups generated from specific Lorenz curves and the parameter values. The last column of the table shows the values of the Theil statistics computed using the analytical expressions derived using the results in Section 3 and the formulae in equations (27), (33) and (38).

**Table: Results from Simulations**

**Chotikapanich Lorenz curve:**  $\eta = \frac{e^{k\pi} - 1}{e^k - 1}$

Parameter Values ( <i>k</i> )	Theil from decile	Theil analytic
1	0.040227152	0.040911353
2	0.149902658	0.151910837
3	0.303075522	0.307247377
4	0.472698353	0.479863262
5	0.639681097	0.650656583
6	0.794193161	0.809783722

**Gupta Lorenz curve:**  $\eta = \pi a^{\pi-1}$

Parameter Values ( <i>a</i> )	Theil from decile	Theil analytic
2	0.057742757	0.058393221
4	0.175846556	0.177790864
6	0.254285285	0.257195897
8	0.310943335	0.314617083
10	0.354771061	0.359078822
12	0.390270074	0.395122705

**Kakwani-Podder Lorenz curve:**  $\eta = \pi * e^{-\beta(1-\pi)}$

Parameter Values ( $\beta$ )	Theil from decile	Theil analytic
1	0.106020473	0.107239445
2	0.295247342	0.298786564
3	0.488006428	0.49461985
4	0.663555448	0.673970622
5	0.81850556	0.833475369
6	0.954457326	0.974747995

Parameter values ranging from one to six have been chosen for Chotikapanich and Kakwani-Podder functional forms to demonstrate the effectiveness of the analytic solutions over a wide range of values. In practice most parameter estimates for the Chotikapanich Lorenz curve lie between two and four, while most parameter estimates for the Kakwani-Podder functional form lie between one and three. Parameter estimates for the Gupta Lorenz curve are typically higher, usually ranging between 4 and 12. All three analytical solutions accurately match the discrete approximations even when using extreme parameter values.

The results presented in the table clearly show that the numerical values of Theil's statistic derived using the discrete approximation formula and those calculated using the analytical expressions derived in the paper are very close. As expected, the values of the Theil statistic from the analytical expressions are all slightly higher than the values from the decile share data. This is because data presented in grouped form ignores the inequality within each group. The analytic solution does not face this drawback. These results confirm the validity and usefulness of the expressions presented in this paper.

## *6. Conclusions*

The paper has provided a formal link between the popular Theil's T statistic and the Lorenz curve which forms the basis for a number of other inequality measures including the Gini coefficient. The paper provides a mathematical expression for the Theil's measure in terms of the Lorenz curve and its first and second order derivatives. The main result of the paper shows that the Lorenz curve can be seen as the basis for most of the commonly used inequality measures. Analytical expressions for the Theil statistics for some specific Lorenz curve forms are derived and results from the simulation experiments clearly demonstrate the validity of the usefulness of the analytical expressions provided in the paper.

## References

Atkinson, A.B., *On the Measurement of Inequality*, Journal of Economic Theory, 2, 1970.

Chotikapanich, D., *A Comparison of Alternative Functional Forms for the Lorenz Curve*, Economics Letters, 41, 1993.

Chotikapanich, D., Rao, D.S. and Tang, K., *Estimating Income Inequality in China Using Grouped Data and the Generalised Beta Distribution*, WIDER research paper, 134, 2006.

Cowell, F.A., *Measuring Inequality (2<sup>nd</sup> edition)*, Harvester Wheatsheaf, Hemel Hempstead, 1995

Gupta, M.R., *Functional Form for Estimating the Lorenz Curve*, Econometrica, 52, 1984.

Kakwani, N.C., *Income Inequality and Poverty: Methods of Estimation and Policy Applications*, Oxford University Press, New York, 1980.

Kakwani, N.C., and Podder: N, *On the Estimation of Lorenz Curves from Grouped Observations*, International Economic Review, 14, 1973.

Milanovic, B., *True world income distribution, 1988 and 1993: First calculation based on household surveys alone*, Economic Journal, 112, 2002.

Sala-i-Martin, X., *The World Distribution of Income (Estimated from Individual Country Distributions)*, NBER Working Paper, 2002.

Theil, H., *Economics and Information Theory*, Amsterdam, North Holland, 1967.