Recent Development of Personal Income Distribution Models: Application on the Case of Rural China

He, Yan
Graduate School of Economics, Kyushu University

https://doi.org/10.15017/25124
Recent Development of Personal Income Distribution Models: Application on the Case of Rural China

Yan He†

Index
1. Introduction
2. A Survey of Income Distribution Function
3. Basic properties and Methodology
  3.1 Properties of Reasonable Mathematical Model
  3.2 The Four-parameter Dagum Distribution
  3.3 Gini Concentration Ratio
  3.4 Methods of Parameter Estimation
4. Application of Dagum Distribution: Household Income Data in Rural China
  4.1 Data Explanation
  4.2 Estimation of Gini Coefficient
  4.3 Lorenz Curve
5. Conclusion

1. Introduction

Inequality has been an issue of concern of social scientists and policy makers in China. China has achieved high economic growth for over 20 years with rising income inequality (Rozelles, 1994; Yao et al., 2004, Wan and Zhang, 2006). To describe the income inequality in China, different inequality index have been employed such as Theil index (Akita, 2003; Cheng and Li, 2006) and Gini coefficient (Yao, 1997, 1999; Wan, 2001). More popular, measure of inequality is the Gini coefficient, which is more intuitive to many people since it is based on the Lorenz curve.

For the case of China, the calculation of Gini coefficient is mostly concentrated on decomposition. Attempt on geographic targeting (Wagstaff, 2005) spatial decomposition has firstly been drawn attention since the coastal areas experienced phenomenal growth due to the open policy in 1978 (Yao, 1997; Yang, 1999; Shorrock and Wan, 2005). Moreover, regression-based decomposition has also been applied to estimate Gini ratio by Zhang and Zhang (2003), Wan (2004), Wan and Zhou (2005) and Deng and Li (2009). The most popular studies devote to inequality decompo-

† Ph.D. Program 2nd Year, Graduate School of Economics, Kyushu University, Japan; Lecturer of Business School in Xiangtan University, Hunan, China. I wish to thank Prof. Saeki for helpful comments. The views and any remaining errors in this paper are solely author’s responsibility. I would like to thank financial support from “Humanities and Social Sciences Planning Fund Supported by Chinese Ministry of Education: The study on the Linkage mechanism between coordination and regional urbanization under the goals of common rich development (No.12YJC790053, 2012~2013 year)”.

sition by the means of subgroups. However, there are some drawbacks on subgroups to calculate Gini coefficient, which is hard to define and interpret surplus term R (Mookherjee and Shorrocks, 1982), although many scholars tried to offer reasonable interpretation for R (Siber, 1989; Yitzhaki and Lerman, 1991; Lambert and Aronson, 1993). Another debate of is that the subgroups decompositions have variable types on weight, such as Rao (1969) demonstrates weights should be population share of subgroup, while Mangahas (1975) debates it is proper to use income share of subgroup for weights. Targeting above issues of estimating Gini coefficient, parameter estimation can solve well. The key point is to choose an effective model to fit the family income data.

There are a great mount of literatures concentrating on model specification from two parameters to five parameters, even more. Two-parameter function is too simple to reflect the impact of economic fluctuation on the size distribution, whereas, five-parameter model is easy to loss simplicity. Thus, the purpose of this paper is to estimate Gini ratio by applying Dagum (1997) distribution (four-parameter model) on Chinese family income data for its various merits comparing with other models. Relative studies such as, Bandourian et al. (2003) find that, in a study utilizing 82 data sets, the log-logistic model (Dagum, 1975) is the best 3-parameter model in no less than 84% of the cases. The Dagum often provides a better fit to income data than the closely related Singh-Maddala distribution, which Kleiber (1996) provides a heuristic explanation.

On the whole, the purpose of this paper is to apply Chinese family income data by employing a 4-parameter model proposed by Dagum (1977a). In China, the statistical data is separated to rural part and urban parts. In this paper, I employ family income data in rural part only. We also apply two methods proposed by Dagum (1977a) for comparison. The rest of the paper will be organized as follows. Section 2 represents the development of parameter estimation, and points out the merits of Dagum distribution. Section 3 discusses the income distribution model and methodology used in this study. Section 4 illustrates applications of the results to rural household survey data in rural China. Section 5 concludes.

2. A Survey of Income Distribution Function

Pareto (1895) first proposed a model of income distribution in the form of a probability density function, which started the exploration of the field of personal income distribution. The apparent attractions of the Pareto distribution evaporate somewhat when one considers its implications for the distribution of income amongst the population as a whole (Clementi et al., 2008). However, empirical studies showed that the Pareto distribution accurately models only high levels of income, but poor in describing the lower end of the distribution since the income distribution is right-skewed and has a fat right-hand tail. As research continued, a variety of probability functions were proposed as suitable in describing the distribution of income by size
using both a combination of known statistical distributions (Nirei and Souma, 2004; Clementi et al., 2005) and parametric functional forms for the distribution of income as a whole.

Two-parameter model is first proposed by Pareto (1895) known as the first Pareto law proved by Champernowne (1953) which demonstrated that under certain assumptions the stationary income distribution will approximate the Pareto distribution irrespectively of the initial distribution, other contributions of mentioned area see Mandelbrot (1961) and Wold and Whittle (1957). Aitchison and Brown (1954) made efforts to complicate the matter with the applicability of the Champernowne’s model. Sargan (1957) built the most complicated model but made it unwieldy and unintelligible (Kleiber and Kotz, 2003). For better describing the observed data, Gibrat (1931) suggested lognormal distribution, which was modified by Kalecki (1945) and examined by Aichinson and Brown (1969) found that the lognormal distribution fit the whole range of income distribution but is quite poor in describing both the upper and lower tails of the actual distribution. To present the satisfactory goodness of fit, Ammon (1895) proposed the gamma distribution, which was applied to fit income data by March (1898) and further promoted by the work from Salem and Mount (1974) and Bordley et al. (1996) for USA and Bartels and van Metelen (1975) for the Netherlands showing that empirical evidence favors the Gamma over the lognormal distribution. Moreover, Rutherford (1955) incorporated birth-death considerations into a Markov model which provides a better fit than the lognormal. In the study of Branchman et al. (1996) shows that the gamma distribution emerges as the best two-parameter model by utilizing German household incomes data. The log-Gompertz distribution appears to be used mainly in income and size distributions and was noticed by Dagum (1980) in this connection, which has been proved to be an excellent two-parameter model by Cummins et al. (1990). Other contributions for two-parameter distribution are made by Bartels and van Metelen (1975) for the Weibull distribution and Kloek and van Dijk (1976) the log t (where t is the Student distribution).

Metcalf (1972) argued that a two-parameter function is too simple to reflect the impact of economic fluctuation on the size distribution. Thus, three-parameter functions are studied by scholars. These include the generalized gamma distribution proposed by Taille (1981) and Amoroso (1924-1925) as well as Beta distribution studied by Thurow (1970) and Kakwani and Podder (1976). Singh and Maddala (1976) and Dagum (1977a) proposed two closely related model which are the members of the Burr family of distributions known as Burr \( \text{VII} \) and Burr \( \text{III} \) in statistics, respectively. The Singh-Maddala distribution is also known as, the beta-P distribution (Mielke and Johnson, 1974), Burr distribution (Hogg and Klugman, 1983, 1984), the Pareto IV distribution (Arnold, 1983), or a generalized log-logistic distribution (El-Saidiet al., 1990). Two-parameter models, including Lomax (1954) distribution, Log-logistic distribution suggested by Fisk (1961) and Para-logistic distribution proposed by Klugman et al. (1998), are the special case of Singh-Maddala distribution. Pareto (1896) also specified the Type II (three-parameter) to be
considered as alternative to the Type I. Furthermore, Vinci (1921) presented Pearson Type V
distribution as a model of income distribution. These distributions allowed for intersecting
Lorenz curves, a phenomenon observed in the data that could not be modeled by any of the
two-parameter distributions considered (Bandourian et al., 2002).

For an accurate description of empirical distributions with an associated measure of income
inequality, four-parameter models were also taken into consideration. Pareto (1897) developed
the Type III (four-parameter) models on the basis of Type I (Pareto, 1895) and Type II (Pareto,
1896). Champernowne (1952) studied the hyperbolic distribution. However, the four-parameter
distributions that were not only very successful in fitting the data, such as the means and Gini
ratio appear to be slightly underestimated (Branchmann et al., 1996), but also included all of the
previously mentioned distributions as special or limiting cases by studying the generalized beta
of the first (GB1) and the second (GB2) kinds (McDonald, 1984). To solve the problems
mentioned, Dagum (1977a) proposed a new model of person income distribution, empirical work
for USA data shows that the results are the best comparing with the Lognormal, the Gamma, and
the Singh-Maddala models.

Furthermore, Five-parameter distributions were presented by Champernowne (1953) and
McDonald and Xu (1995). As is mentioned above, two-parameter models usually failed to be
empirically corroborated, while five-parameter models are prone to loss simplicity. For the
merit of the model proposed by Dagum (1977a), I will apply to Chinese data to estimate the
inequality of China for the time period of 1985 to 2010.

3. Basic properties and Methodology

3.1 Properties of Reasonable Mathematical Model

The specified model may be governed by its capacity to account fairly well to a set of economic,
econometric, stochastic and mathematical properties. Dagum (1977a) proposed a personal in-
come distribution model which supports intersecting Lorenz Curves and convergence to the weak
form of Pareto distribution\(^1\) and satisfies all the properties as follows.

Firstly, requirement for the parameter set. The mathematical form of the distribution func-
tion, which must be simple from a technical point of view and fundamental from the point of view
of model-building, can be derived from an elementary set of logic-empirical postulates or
assumption. For the setting of unknown parameters, the principle is to make use of the smallest
possible number of parameters for adequate and meaningful representation. The number should
not be either too simple to empirically corroborated with data or too complicated to loss

\(^1\) For details about the interpretation of intersecting Lorenz Curves and convergence to the Pareto distribution,
please refer to Dagum (1977a)
simplicity. Thus, three- or possible four- parameter function is taking into consideration. Furthermore, all the parameters should have well-defined economic meaning and simple and efficient of parameter estimation, which is always an advantage from the point of view of computer cost and the acceptance of model in applied economics.

Secondly, concerns on model flexibility. Specified model solves the problem of the model being identified and the actual observation, and provides a good fit of whole range of the distribution. An ideal model is able to deal with a positive, and not predetermined, minimum income without truncating the distribution. Especially, the existence of negative and nil income, which strongly restricts the descriptive power, is considered. Furthermore, the properties considered, such as the shape of distribution though changes in parameter values, unimodal and strictly decreasing (non-modal) income distribution, are also important.

Thirdly, conditions on the parameter estimation. Cumulative distribution function approach was chosen by Dagum to overcome the assumption of equi-distribution within each interval of income required when dealing with the method of parameter estimation. Dagum distribution gives the explicit solution for Lorenz curve and Gini concentration ratio, which allows computing Gini coefficient directly and verifying whether it is proposed as a goodness of fit test.

3.2 The Four-parameter Dagum Distribution

He motivates his model from the empirical observation that the income elasticity of the cumulative distribution function (CDF) \( F \) of income is a decreasing and bounded function of \( F \). Starting from the differential equation:

\[
\frac{d \log[F(x)-a]}{d \log x} = \beta \delta \left[ 1 - \left( \frac{F-a}{1-a} \right)^{1/\beta} \right]
\]

where \( x > 0 \) if \( 0 < a < 1 \) and \( x > x_0 > 0 \); \( F(x_0) = 0 \), if \( a < 0 \), subject to \( \beta > 0 \), \( \delta < 1 \), and \( \delta \beta > 1 \).

The solution of function (1) can get cumulative distribution function (c.d.f):

\[
F(x) = a + \frac{1-a}{(1+x^{-\delta})^\beta}, \lambda > 0
\]

where \( \lambda \) is a scale parameter and strictly positive because it is the antilog of the constant of integration. \( a \), \( \beta \) and \( \delta \) are dimensionless parameters. \( a \) is an inequality parameter and \( \beta \) and \( \delta \) can be called equality parameters because the Gini ratio is an increasing function of the former and a decreasing function of the latter.

3.3 Gini Concentration Ratio

Cumulative distribution function (2) associated with the Lorenz curve argued by Dagum (1977b) can be noted as:
\[ L(y) = \frac{B(y^{1/a}, \beta + \frac{1}{\delta}, 1 - \frac{1}{\delta})}{B(\beta + \frac{1}{\delta}, 1 - \frac{1}{\delta})}, \beta \delta > 1, 0 \leq a < 1 \] (3)

where
\[ y = \frac{F(x) - a}{1 - a}, y \in [0, 1] \]

and its corresponding Gini concentration ratio is:
\[ G = \frac{(2a - 1)(1 - a)\Gamma(\beta)\Gamma(2\beta + \frac{1}{\delta})}{\Gamma(2\beta)\Gamma(\beta + \frac{1}{\delta})} \] (4)

where \( \Gamma(\cdot) \) denotes the complete Gamma function.

3.4 Methods of Parameter Estimation

In order to estimate the parameters of the above model, there are five methods proposed, including iterative method I, II and III; unconstrained function minimization and the method of maximum likelihood. For comparison, this paper will choose iterative method I and maximum likelihood to apply on Chinese data.

**Iterative method I**: According to the transformation of equation (2), we can get:
\[ \log F(x) = -\beta \log(1 + \lambda x^{-\delta}) + \sum_{k=1}^{m} a_{k} \frac{1 - F(x)}{kF(x)} \] (5)

Equation (10) can be approximated by the following linear form:
\[ \log F(x) \approx a \frac{1 - F(x)}{F(x)} - \beta \log(1 + \lambda x^{-\delta}) \] (6)

To estimate the parameter vector \((a, \beta, \lambda, \delta)\) in equation (2), start with an initial value of \(a, \beta, \lambda\) and \(\delta\) by nonlinear least square using equation (6) to search for convergence by the iterative procedure.

**The method of maximum likelihood (ML)**: According to Wold (1961,1963), assume \(F(x)\) in equation (2) be an unbiased predictor of the sample realization \(y(x)\). Let \(\varepsilon\) is the purely random variable and normally distributed, Hence,
\[ y(x) = F(x) + \varepsilon, \varepsilon \sim N(0, \sigma^2) \] (7)

The log of the likelihood equation is
\[ \log L = -\frac{n}{2} \log 2\pi \frac{n}{2} \log \sigma^2 - \frac{1}{2} \sigma^2 + \sum (y(x) - a - (1 - a)(1 + \lambda x^{-\delta})^2) \] (8)

The ML estimator \(\hat{\mu}\) of \(\mu\) is obtained by solving the likelihood equation \(\frac{d \log L}{d \mu} = 0\), where \(\mu = (a, \beta, \lambda, \delta)\).
4. Application of Dagum Distribution: Household Income Data in Rural China

4.1 Data Explanation

I will apply Dagum distribution on family income data of rural China for the time period from 1985 to 2010. Income data is gathered from “China Statistical Yearbook” and “China Rural Household Survey Yearbook”, which grouped by the percentage of households. Hence, it is need to transfer household proportion to population proportion. To simplify the mathematical expression, the variables defined are as follows.

**PP**: Population proportion

**SHE**: The surveyed households of each group

**TSH**: Total surveyed households

**APE**: Average population in each group

**TAP**: Average population in whole surveyed households

Household population is transferred to individual population proportion by the function:

$$PP_i = \text{SHE}_i \times \text{APE}_i / \sum_{n=1}^{20} (\text{SHE}_i \times \text{APE}_i) = \text{SHE}_i \times \text{APE}_i / \text{TSH} \times \text{TAP}$$  \hspace{1cm} (9)

where $i$ denotes income class. The data structure in Yearbook is arranged by “Percentage of Households Grouped by per Capita Annual Net Income (%)}”, which we denote as PH. Thus, equation (9) can be expressed as:

$$PP_i = \text{SHE}_i \times \text{APE}_i / \text{TAP}$$  \hspace{1cm} (10)

Furthermore, we take mean income of net income as income variable. For comparison to native studies, the mean value of each income range will also be used.

4.2 Estimation of Gini Coefficient

The methods of Iterative I and Maximum Likelihood will be used to estimate the parameters and calculate Gini ratio. RATS software of version 7 is employed to compute the parameters. The sum of squared residuals of the observed cumulative distribution function ($F(x)$) from their corresponding predicted probabilities are denoted by:

$$\sum \varepsilon^2 = \sum \left( F(x) - \hat{F}(x) \right)^2$$  \hspace{1cm} (11)

The estimated parameter vector ($\alpha$, $\beta$, $\lambda$, $\delta$) by the method of Iterative I lists in table 1. The values of all parameters are positive and significant. $\beta$ and $\delta$ are significant at the level of 1%. $\lambda$ and $\alpha$ are at least in the significant level of 10%. It also can be seen that the sums of squared residuals show a satisfied goodness fit.

The last column reports the Gini concentration ratio applying the formula deduced from model (2) and presented in (4). Gini ratio in rural China shows an increasing trend from 0.25151 in 1985
to 0.40117 in 2009 which was over the international warning level of 0.4. In 2010, Gini coefficient is slow down. Whatever, high Gini coefficient shows high income inequality and serious social problem in rural China.

Of special interest are methods fitting to the same data, with an eye on relative performance. To avoid the approximation in equation (6), Table 2 represents the estimation results of maximum likelihood. Comparing with Iterative I, the estimated values of all parameters in Table 2 are significant at the level of 1% except parameter \( \alpha \). Variance is approximate to zero, which shows an excellent goodness of fit.

Furthermore, Gini ratio in Table 2 is higher than that of in Table 1 correspondingly, which is can be seen in Figure 1. The line of maximum likelihood is above iterative I. The possible reason accounts for approximate (6). Gini ratio estimated by maximum likelihood is more reliable, although the results of two methods show the increasing trend.

Native studies also estimated Gini coefficient by different methods. For example, Hong (2008) developed a new decomposition method to estimate Gini coefficient of China. With increasing trend, Gini coefficient estimated by Hong (2008) is lower than either of the methods proposed by Dagum (1977a) except the year of 1990, which demonstrate that the Gini coefficient is probably underestimated by traditional calculation method.

Table 1 Method: Iterative I

<table>
<thead>
<tr>
<th>Year</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \lambda )</th>
<th>( \delta )</th>
<th>( \sum e^2 )</th>
<th>No. of intervals</th>
<th>Gini Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>0.045565***</td>
<td>1.446203***</td>
<td>508440452.6205***</td>
<td>2.604936***</td>
<td>0.049</td>
<td>20</td>
<td>0.38891</td>
</tr>
<tr>
<td>2009</td>
<td>0.011531*</td>
<td>0.933395**</td>
<td>589929808.5638**</td>
<td>2.57269***</td>
<td>0.296</td>
<td>20</td>
<td>0.40117</td>
</tr>
<tr>
<td>2005</td>
<td>0.019391**</td>
<td>1.108717***</td>
<td>507402433.8652**</td>
<td>2.562319***</td>
<td>0.652</td>
<td>20</td>
<td>0.39461</td>
</tr>
<tr>
<td>2004</td>
<td>0.011069***</td>
<td>1.194235***</td>
<td>332116907.2456**</td>
<td>2.568267***</td>
<td>0.431</td>
<td>20</td>
<td>0.38357</td>
</tr>
<tr>
<td>2002</td>
<td>0.01384***</td>
<td>1.21635***</td>
<td>241181256.4741*</td>
<td>2.595083***</td>
<td>0.230</td>
<td>20</td>
<td>0.38019</td>
</tr>
<tr>
<td>2000</td>
<td>0.007069***</td>
<td>1.082874***</td>
<td>243540044.8901**</td>
<td>2.603951***</td>
<td>0.064</td>
<td>20</td>
<td>0.38251</td>
</tr>
<tr>
<td>1999</td>
<td>0.004360***</td>
<td>1.222829***</td>
<td>278013552.4498**</td>
<td>2.643996***</td>
<td>0.053</td>
<td>20</td>
<td>0.36688</td>
</tr>
<tr>
<td>1998</td>
<td>0.005991***</td>
<td>1.273749***</td>
<td>331088981.7561***</td>
<td>2.683727***</td>
<td>0.041</td>
<td>20</td>
<td>0.35990</td>
</tr>
<tr>
<td>1995</td>
<td>0.007274***</td>
<td>1.072681***</td>
<td>562480884.9372***</td>
<td>2.859753***</td>
<td>0.024</td>
<td>20</td>
<td>0.34938</td>
</tr>
<tr>
<td>1990</td>
<td>0.005364***</td>
<td>1.228693***</td>
<td>422389682.4205***</td>
<td>3.631742***</td>
<td>0.129</td>
<td>12</td>
<td>0.26690</td>
</tr>
<tr>
<td>1985</td>
<td>0.021245***</td>
<td>1.338987***</td>
<td>3526068765.6278***</td>
<td>3.980777***</td>
<td>0.026</td>
<td>12</td>
<td>0.25151</td>
</tr>
</tbody>
</table>

Note:***, ** and * represent at the significant level of 1%, 5% and 10%, respectively.
Table 2  Method: Maximum likelihood

<table>
<thead>
<tr>
<th>Year</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \lambda )</th>
<th>( \delta )</th>
<th>VAR</th>
<th>No. of intervals</th>
<th>Gini Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>0.05418***</td>
<td>42.82797***</td>
<td>613891.127***</td>
<td>2.2247***</td>
<td>0.00075***</td>
<td>20</td>
<td>0.40182</td>
</tr>
<tr>
<td>2009</td>
<td>-0.110558***</td>
<td>0.357861***</td>
<td>3214082008.5156***</td>
<td>2.5848***</td>
<td>0.003127***</td>
<td>20</td>
<td>0.45142</td>
</tr>
<tr>
<td>2005</td>
<td>0.005896**</td>
<td>1.105807***</td>
<td>101718795.0429***</td>
<td>2.3644***</td>
<td>0.000069***</td>
<td>20</td>
<td>0.41879</td>
</tr>
<tr>
<td>2004</td>
<td>-0.022947***</td>
<td>0.741795***</td>
<td>1077813740.4201***</td>
<td>2.6314***</td>
<td>0.000122***</td>
<td>20</td>
<td>0.39248</td>
</tr>
<tr>
<td>2002</td>
<td>-0.015134***</td>
<td>0.612957***</td>
<td>7588416842.9366***</td>
<td>2.9048***</td>
<td>0.000230***</td>
<td>20</td>
<td>0.38008</td>
</tr>
<tr>
<td>2000</td>
<td>0.002252*</td>
<td>1.107039***</td>
<td>88424121.0815***</td>
<td>2.4690***</td>
<td>0.000081***</td>
<td>20</td>
<td>0.38579</td>
</tr>
<tr>
<td>1999</td>
<td>0.014402***</td>
<td>2.177649***</td>
<td>11318825.7531***</td>
<td>2.3067***</td>
<td>0.000157***</td>
<td>20</td>
<td>0.39796</td>
</tr>
<tr>
<td>1998</td>
<td>0.005275*</td>
<td>1.265124***</td>
<td>214724501.5825***</td>
<td>2.6185***</td>
<td>0.000062***</td>
<td>20</td>
<td>0.36895</td>
</tr>
<tr>
<td>1995</td>
<td>0.004793**</td>
<td>1.127685***</td>
<td>145360478.6224***</td>
<td>2.6746***</td>
<td>0.000121***</td>
<td>20</td>
<td>0.36824</td>
</tr>
<tr>
<td>1990</td>
<td>-0.0128***</td>
<td>0.584200***</td>
<td>1.2334e+012***</td>
<td>4.3199***</td>
<td>0.000030***</td>
<td>12</td>
<td>0.26784</td>
</tr>
<tr>
<td>1985</td>
<td>0.009374*</td>
<td>1.277819***</td>
<td>1545345040.9909***</td>
<td>3.8216***</td>
<td>0.001768***</td>
<td>12</td>
<td>0.25467</td>
</tr>
</tbody>
</table>

Note:***, ** and * represent at the significant level of 1%, 5% and 10%, respectively.

Figure 1  The comparison of Gini Coefficient between Iterative I and Maximum Likelihood (ML)

Figure 2  The comparison of Gini Coefficient between Dagum (1977a) and Hong (2008)
4.3 Lorenz Curve

Lorenz curve offer an intuitive path to identify inequality. We take Excel to draw Lorenz curve by original data. To investigate the trend of change, take the years of 1990, 2000, 1995, 2000, 2005 and 2010 for example.

At the beginning of drawing the graphs, we calculate cumulative proportion of income. It can be seen from Figure 3 to Figure 8 that the shape of the Lorenz Curves is changing through the time period. The situation of income inequality is worsening. In 1990, approximate 50% of the population own 20% of the income. High income class, about 1% of the population takes up 10% of the income which caused by the reform of open policy in 1978. Income inequality began to increase. In 2000, the condition of low income class improved, almost 34% of the population has 20% of the income. For high income class, about 10% of the population has 30% of the income.

The following year, income inequality increase gradually for the labor flows from rural areas into urban areas. In 2005, about 18% of the population has 21% of the income for the lower class. A series of policies prone to rural China and surplus labor force forcing farmers flowing back to rural areas, the situation of income inequality become well in 2009. In 2010, the situation is better for a series policies bias to develop rural areas. Overall, the income inequality in China is an increasing trend. For all time range, high income class owns the majority of the income. Income inequality with increasing trend in rural China is an issue of policy maker.
Figure 4  Lorenz curves in 1995

Figure 5  Lorenz curves in 1995
Lorenz Curve of 2000

Figure 6  Lorenz curves in 2000

Lorenz Curve of 2005

Figure 7  Lorenz curves in 2005
5. Conclusion

This Chapter has provided a brief introduction to the development of models for income distribution and applies Dagum (1977a) distributions to household data in rural China for the time period of 1985 to 2010. Two methods including Iterative I and maximum likelihood are employed to estimate the parameters of Dagum model. I conclude that Dagum distribution fits household data well in the case of rural China. Maximum likelihood prefers to offer more reliable results for our study since it avoids the approximate problem in Iterative I method which will underestimate the inequality level. Moreover, comparing Dagum distribution with native study, we found that Gini coefficient calculated by Dagum distribution is higher than native estimation by traditional paths. That is, Gini coefficient might be underestimated by the native estimation. This study shows that Gini coefficient is mainly decided by the method applied. Various models offer many kinds of results. It is necessary to choose the suitable model to fit different types of data.

References


Recent Development of Personal Income Distribution Models: Application on the Case of Rural China

publication, Canada.


