The University of Manchester
Brooks World Poverty Institute

MANCHESTER 1824

The trend of the Gini coefficient of China

Jiandong Chen[1]
Dai Dai[2]
Ming Pu[3]
Wenxuan Hou[4]
Qiaobin Feng[5]

[1] Southwestern University of Finance and Economics, Chengdu, PR China
cjd9754@yahoo.com

[2] Southwestern University of Finance and Economics, Chengdu, PR China
postdd@swufe.edu.cn

[3] Southwestern University of Finance and Economics, Chengdu, PR China
pu_ming@hotmail.com

[4] Durham University, UK
wenxuanhou@126.com

[5] Southwestern University of Finance and Economics, Chengdu, PR China
qbfeng@hotmail.com

January 2010

BWPI Working Paper 109

Brooks World Poverty Institute
ISBN : 978-1-907247-08-8

Creating and sharing knowledge to help end poverty

www.manchester.ac.uk/bwpi

Abstract

A literature review indicates that the main problem in calculating the Gini coefficient of Chinese residents' income is the shortcomings of the data sources. Though many studies have tried to overcome these limitations through decomposing the nationwide Gini ratio by urban and rural areas, the final results have been underestimated, due to the overlap term or residual in the decomposition. This paper analyses the effects of the overlap term on calculating the overall Gini coefficient through a statistical approach, and estimates Chinese Gini ratios since economic reform and open door policies were adopted. Based on decomposing the Chinese Gini coefficient from 1978 to 2006, the authors find that the key factor of income inequality comes from income disparity between rural and urban inhabitants. The authors investigate the features of this income inequality between rural and urban areas. Furthermore, statistical approaches are employed to evaluate the effects of the development of urbanisation and rural-to-urban average income on the income inequality of the whole nation. The results show that accelerating the pace of urbanisation is the key issue to improving Chinese income disparity. On the basis of the above analysis, the paper proposes related policies for policy-makers.[1]

Keywords: Chinese Gini coefficient; Decomposition of Gini coefficient; Urbanisation

Jiandong Chen[2] is Associate Professor, South Western University of Finance and Economics (SWUFE), Chengdu, China.

Dai Dai is a Lecturer at South Western University of Finance and Economics (SWUFE), Chengdu, China.

Ming Pu is Associate Professor, South Western University of Finance and Economics (SWUFE), Chengdu, China.

Wenxuan Hou is Lecturer in Finance, Durham Business School, Durham University, UK.

Qiaobin Feng is Professor, South Western University of Finance and Economics (SWUFE), Chengdu, China.

[1] This study is subsidised by the project of '211' three sub-discipline construction with subject of Public Finance of Southwest University of Finance and Economics.

[2] Corresponding author.

1. Introduction

There are various means to measure income inequality, among them the Theil index, coefficient of variation, Kuznets index, the Gini coefficient, etc. However, the Lorenz curve and the Gini coefficient are most commonly used (Sloman, 2000). The Gini coefficient is the most important index for measuring or estimating income inequality (Sen, 1997; Champernowne and Cowell, 1998). In 1912 the Italian statistician, Corrado Gini, published *Variability and Mutability*, in which he proposed a method to measure inequality. This method gradually evolved into the well-known Gini coefficient (Shi Li, 2002).

There are many disagreements on how to calculate the Gini coefficient for Chinese residents. According to our literature review, there are about 20 different estimations on Chinese Gini ratios. Nothing can illustrate this issue more than different estimations of the Gini ratio for the year 1995. Zongsheng Chen (1999) calculated it as 0.365. However, in 2002, he and Yunbo Zhou used two other methods to calculate it, with the result of 0.38392 and 0.41914, respectively. The results from Yonghong Cheng (2006, 2007), and Chen, Hou and Jin (2008) were 0.4169, and 0.3934, respectively. Shujian Xiang (1998) and Dan Huang and Youmin Xi (1999) obtained results of 0.3515 and 0.328, respectively. Renwei Zhao, Shi Li, Riskin (1999) obtained 0.445. The highest one, 0.452, was provided by Khan and Riskin (2001). This value is nine percent higher than the 0.415 obtained by Ravallion and Chen (2007). It was almost twice the lowest value. Although the Rural and Urban Socio-Economic Survey Teams under the National Bureau of Statistics estimated the intra-rural and the intra-urban Gini coefficient, respectively, since 1978, and their results are widely quoted, there are still disagreements on how to calculate the nationwide Gini coefficient since economic reform and open door policies were adopted. The Gini coefficient provides basic statistical data for analysing income inequality in China. Therefore, such variation in approaching it limits the progress of studies on China's income inequality.

This paper investigates the influences of the income overlap part on the nationwide Gini coefficient. Then we present a new approach to estimating the Chinese Gini ratio from 1978 to 2006, which avoids the shortcomings of current data sources. In line with the results, the authors further probe the trend of Chinese income disparity. The decomposed nationwide Gini ratio indicates that the dominant issue in Chinese income inequality is income disparity between rural and urban areas. Meanwhile, we find that urbanisation is the key factor in narrowing the Chinese rural–urban income gap. Speeding up urbanisation will thus help to improve Chinese income disparity.

The framework of this paper is arranged as follows: Part 2 analyses the shortcomings in calculation of the Chinese Gini coefficient, based on a literature review; Part 3 concentrates on examining how the overlap terms influence estimation of the Gini ratio;

Part 4 introduces how to estimate Chinese Gini ratios, decomposing the Gini ratio by cities and countryside; Part 5 focuses on how the income gap between the cities and the countryside affects the income inequality of the nation, focusing particularly on the influence of urbanisation and the urban-to-rural income ratio on nationwide income inequality; and the final part presents the conclusion and suggestions for policy makers.

2. Shortcomings of current estimations on the Chinese Gini coefficient

The ideal statistical source for calculating the Gini coefficient would be original household survey data. One can easily and accurately calculate the Gini ratio through certain software[3] based on the formula (1):

$$G = \frac{1}{2n^2 u} \sum_i \sum_j \left| y_i - y_j \right| \quad (1)$$

In formula (1), *n* stands for the population (or the family count); *u* is the average income; y_i and y_j stand for the income of *i* family and *j* family, respectively. However, the publicised Chinese income data are not the original ones any more. The original ones were divided into groups based on income level. For example, the *China Statistical Yearbook* only divided urban resident samples, which covered over 60,000 households, into seven groups. Since there are only a few groups, accuracy in calculating the Gini ratio cannot be assured. Based on a literature review, we argue that the shortcomings of estimating the Chinese Gini ratio come mainly from the aspects outlined.below.

(1) Based on *China Statistical Yearbook* and similar data sources

Several statistical yearbooks have been published, such as *China Statistical Yearbook*, *Yearbook of Urban Living and Price Index*, *Yearbook of the Rural Household Survey*,[4] and *Statistical Yearbook of Chinese Price Index* and *Urban Household Survey of Income and Expenditure*. The last two group the sampled households in a similar way to the *China Statistical Yearbook*. Although *Yearbook of Urban Living and Price Index* provides more income groups than *China Statistical Yearbook*,[5] it started to be officially published only in 2006.

---

[3] Such software includes 'DAD' developed by LAVAL, which is able to calculate the index for extreme income inequality. Shi Li (2002) also pointed out that one can use software specialising in calculating the inequality index, such as INEQ or Stata, to calculate the Gini ratio. All calculations in this study are done by the Matlab.

[4] *Yearbook of the Rural Household Survey* was first published in 1992. However, it was not until 2000 that the yearbook started to be published consecutively.

[5] *Yearbook of Urban Living and Price Index* categorised all samples into 20 groups based on their income.

China Statistical Yearbook is the most important data resource for calculating the Chinese Gini ratio, especially when no household survey data were available. Zongsheng Chen and Yunbo Zhou (Zongsheng Chen, 1999; Yunbo Zhou and Zongsheng Chen, 2002) calculated the Gini coefficient based on this yearbook. However, the income provided in this yearbook has been greatly doubted. Khan and Riskin (2001) commented that the statistical data in the yearbook were too aggregated to prohibit carefully analysis of income inequality. Fang, Zhang, and Fan (2002) also believed that the aggregated data ignored the income disparity within each group, and there they are not accurate enough. In 1996, *China Statistics Yearbook* stated that those with income more than RMB2000 constituted 38.4 percent among all the households in rural areas. However, the *Yearbook* did not divide such households into subgroups any more. In 2007, the *Yearbook* stated that those with income more than RMB5000 made up 30.94 percent of all the households in rural areas. Yet, after almost ten years, it still did not provide any more subgroups.

Another problem in calculating the Chinese Gini coefficient is that *China Statistical Yearbook* provides two categorical samples – one from urban areas, and the other from rural areas. It does not provide integrated samples for all. In order to solve this problem, Zongsheng Chen and Yunbo Zhou (2002) proposed a statistical approach. This method first adds different weights to rural and urban samples, based on the actual population ratios, and then combines them together. There is nothing wrong with the method itself. However, it does not solve the problems of *China Statistical Yearbook*, including that the groups were over-concentrated, the differences within each group are ignored, etc.

Khan and Riskin (2001) have different opinions on the standard of categorising the residents' income in China. They argued that the standard was different from the international one. According to Shi Li and Chuliang Luo (Shi Li and Chuliang Luo, 2007; Shi Li, 2003), the hidden subsidies of urban residence are far higher than those of rural residence. Rural residents do not have such benefits, including houses, medical insurance, pension insurance and unemployment insurance. If all these hidden benefits are taken into consideration, the income ratio between the cities and the countryside will increase by one-third. Therefore, the actual income gap between cities and countryside is definitely larger than the one provided by the *Yearbook*.

Also, due to the change in statistical methods, there are obvious inconsistencies and inaccuracies with the data provided. Two variables, income and population, are essential to calculate the Gini coefficient. However, the *Yearbook* does not provide consistent statistics standards for income and population.

Before 1991, the income level was set as the average income rather than average disposable income. However, the National Bureau of Statistics has been using the average disposable income as its statistics source since 1995. Khan and Riskin (2001) had different opinions on the standard of categorising the income of citizens in China.

They thought the standard was different from the international one. What is more, the *China Statistical Yearbook* has some noticeable mistakes. In 1994, the packet number of the income of rural residents was over 100 percent more than the total sample numbers for 1985 and 1991. The data of 1991 provided in 1993 and 1995 were also different from each other,[6] and different from those provided in 1994.

Moreover, the standards vary from year to year. Before 1982, the statistics were based on registered residences; from 1982 to 1989, the numbers were derived from the third and the fourth census; from 1990 to 2000, the data were derived from the fifth census in 2005; after 2001, the data were from the sample surveys. Also, the fifth census regarded city residences as whoever had lived in a city for more than six months, while the fourth census regarded city dwellers as people who had lived in a city for more than a year. Nevertheless, the household registration system in China states that city residents are those whose registered residences are in cities, regardless of where they may live. This standard is very different from the ones used by the *Statistical Yearbook*. There will be registered city residents who may live in the countryside for more than a year and there are many countryside people who do not have city residence but reside in cities for long periods. Therefore, statistics vary greatly, due to the different standards used year by year. For example, in 1998, the percentage of the urban population was 33.35 percent, based on the fifth census, 30.4 percent, based on the fourth census, and 24.7 percent, based on the registration system.

Usually, people do not notice the difference in approaches of census between China and other countries in the world. For instance, households are required to record their total annual income and expenses. In contrast, many other countries use one week's, two weeks' or one month's income statements as the basis, and multiply it by 52, 26 or 12 to get the data for a whole year (Gibson, Huang and Rozelle, 2001). Comparatively speaking, a one-year model will decrease fluctuations of income. For example, one month's shortage can be compensated by another month's windfall. Therefore, the Chinese Gini ratio is easily underestimated compared with many other nations (Deaton, 1995).

Despite all the limitations, *China Statistical Yearbook* is the only data resource available to calculate the Chinese Gini index since the economic reform (Fang, Zhang and Fan, 2002). Most other resources can be only dated to 2000.

(2) The Gini coefficient based on other data sources

Using the adjusted data from surveys of city and rural residences in 1988 and 1995,

---

[6] See *China Statistical Yearbook* (1993: 311); *China Statistical Yearbook* (1994: 276); and *China Statistical Yearbook* (1995: 278).

Khan and Riskin (2001), Renwei Zhao, Shi Li and Riskin (1999) all concluded that the income inequality has been rising in China. Their conclusion attracted wide attention. The sample of 1998 included 10,258 residences from 28 provinces, while the one in 1995 included 7,998 residences from 19 provinces. However, the corresponding data from the National Statistics Bureau were 51,352 and 34,739 residences, respectively. Although the sample data were much smaller than those of the National Bureau of Statistics, the estimations were much more accurate than other results. Nevertheless, it is hard to estimate the long-term fluctuations in income inequality with just two years' data. State Council Research Group (1997) also provided a few years' estimations of the nationwide Gini ratio. However, it is hard to accurately reflect the general trend of income inequality in China with a few years' data. If we can access the original household survey data, then we will calculate the Gini coefficient fast and accurately with formula (1). [7] However, we are far from reaching this ideal status with the current data sources, and there are still many technical difficulties to overcome.

(3) Different methods

There are many shortcomings with the published data, and one cannot obtain the original data source used by the National Bureau of Statistics. This situation has forced researchers to find another way to solve the problem. Currently, many researchers decompose the Gini coefficient to calculate the Chinese Gini ratio.

The decomposition of the Gini index is not a new topic in econometrics. It is well known how to decompose the Gini index from different income resources. Nevertheless, it is still difficult to decompose the Gini index from different groups.[8] Bhattacharya and Mahalanobis (1967) were the first to start the research. Pyatt (1976) divided the Gini index based on different income levels according to Game theory. Based on matrix and covariance, Mookherjee and Shorrocks (1982), Shorrocks (1984), Lambert and Aronson (1993), as well as Cowell (2000), also tried to develop new methods to decompose the Gini ratio. However, their methods were not easily understood. Yao (1997) provided a comparatively simpler way to decompose the Gini ratio. In brief, the Gini ratio of the whole population could be decomposed into the following (see Yao, 1997):

$$G = G_g + \sum_{i}^{n} P_i I_i G_i + G(f) \quad (2)$$

Here G_g stands for the Gini ratios among different groups; G_i stands for the Gini ratio within the i^{th} group; the proportion of the income of group i^{th} to the total income is I_i; the proportion of the population of i^{th} group to the total population is P_i; the value of $G(f)$ depends on how much the overlap there is among the different groups. If there is no overlap, the $G(f)$ would be 0 (Mookherjee and Shorrocks,1982; Shorrocks and Wan,

[7] Consider the methods of statistics when doing the international comparison.
[8] Khan and Riskin (2001: 28-29) did not regard it as appropriate to decompose the Gini coefficient based on different groups.

2005). Milanovic (2002) calculated and indicated that the Gini coefficient all over the world in 1993 was 0.578, among which *G* (*f*) was 0.068. This formula explains why the Gini ratio is underestimated based on the statistical yearbooks. In *China Statistical Yearbook*, the incomes of city and rural residents were ranked from low to high, therefore, the *G* (*f*) within either the city group or the rural group was 0. However, when using the data from *China Statistical Yearbook*, there is no way to calculate the Gini ratio within each group, which equals to ignoring the $\sum_{i}^{n} P_i I_i G_i$ of formula (2), and thus the final results are led to be underestimated.

Suppose the Gini coefficient ($G^{'}$) of the whole country can be divided into three parts: the intra-rural Gini ratio (G_r), the intra-urban Gini ratio (G_u), and the Gini ratio between the urban and rural areas (G_{ur}). Then we can get (Chen, Hou and Jin, 2008):

$G = G_{ur} + \delta G_u + \beta G_r$ (3)

δ and β stand for the results of the population proportion of the urban and rural areas, respectively (P_u□P_r□, multiplied by the income proportion of the urban and rural areas respectively (I_u□I_r□, which are $\delta = I_u P_u$ ($\beta = I_r P_r$). The decomposed nationwide Gini ratio includes three segments: the intra-rural Gini ratio, the intra-urban Gini ratio, and the Gini ratio between the urban and rural areas. Their relevant coefficients are δ□ β and 1, respectively. This formula clearly shows the contributions of each part to the nationwide Gini ratio. If there is overlap between the income of urban areas and rural areas, for example, some urban residences' incomes are lower than the higher incomes of the rural residences. If so, formula□3□becomes formula□4□

$G = G_{ur} + \delta G_u + \beta G_r + G_0$ (4)

Here G_0 comes from the overlap term of rural and urban income distribution. Theoretically, if there is one rural residence's income that is is higher than that of an urban residence, then G_0>0. Obviously, the use of formula (3) will underestimate the nationwide Gini ratio if there is overlap part between rural and urban income distribution. To take an extreme example, when the population and income of a rural area is totally equal to that of urban area, the Gini index from formula (3) will be underestimated by 50 percent. Therefore, the value of G_0 is 0□0.5 *G.* As a matter of fact, formula (4) represents a specific case of formula (2). When Chen, Hou and Jin (2008) calculated the Chinese Gini ratio, they failed to consider the influence of G_0, and thus the result was underestimated.

Currently, the domestic literature, such as Zongsheng Chen and Yunbo Zhou (2002), and Shujian Xiang (1998), and foreign documents such as Yang (1999) and Sundrum (1990), all showed the calculation of the nationwide Gini ratio as the following formula:

$G = P_r^2(\frac{Y_r}{Y})G_r + P_u^2(\frac{Y_u}{Y})G_u + \frac{P_u P_r \left|Y_u - Y_r\right|}{Y} + G_0$ (5)

In formula□5□, Y_u□Y_r and Y are the annual average income of urban residences, rural residences and whole population, respectively. Zongsheng Chen and Yunbo Zhou (2002) and Shujian Xiang (1998) used this formula without considering the income overlap between urban areas and rural areas, thus G_0 was ignored, which resulted in an underestimation.

Actually, $\frac{P_u P_r \left|Y_u - Y_r\right|}{Y}$ is the Gini ratio between rural and urban areas (derivations omitted). Since the value of Y_u and Y_r are different in different societies, we take its absolute value. Although formulas (4) and (5) look different, they are essentially the same after transformation (derivation omitted). However, formula (4) reflects the contributions from the intra-rural Gini ratio (G_r), the intra-urban Gini ratio (G_u), and the Gini ratio between urban and rural areas (G_{ur}) in a more visual way than formula (5). Based on formula (4), the contributions from the intra-rural Gini ratio (G_r), the intra-urban Gini ratio (G_u), and the Gini ratio between urban and rural areas (G_{ur}) are $\beta G_r/G$□$\delta G_u/G$ and G_{ur}/G, respectively.

With new methods, Yonghong Chen (2006, 2007) tried to solve this problem – the underestimation due to the overlap between high rural income group and low urban income group. This method needs to presume the income distribution function of urban and rural residences. However, there are not many samples in the statistical yearbooks. Such limited groups can hardly derive a reliable function of Chinese income distribution as a whole (Chunlei Wang, 2007; Suxin Huang 2007). According to Chen's result, the contribution of the Gini ratio between urban and rural areas (G_{ur}/G%) in 1990 was 22.24 percent. This number is quite different with other studies, such as the one conducted by Zongsheng Chen and Yunbo Zhou (2002). Their results are close to 53 percent. Yifu Lin, Fang Cai and Zhou Li (1998) used Theil Entropy to examine the income disparity among rural areas, among urban areas and between urban and rural areas. They found that the difference between urban and rural areas is the most influential on the overall effect. It almost always stays about 50 percent. The study conducted by Houkai Wei (1996) showed that such contribution from 1985 to 1995 averages about 51 percent. Khan and Riskin (2001) stated further that the Chinese Gini ratio is higher than that of rural areas and that of urban areas, which demonstrated the importance of the income gap between the urban and rural areas. Also, similar results were concluded by Guanghua Wan (2004), Jiandong Chen, Shengwu Jin, Xinyue Tang (2005), and Shi Li, Sicular and Gustafson (2008).

Xingjian Hong (2008) used the method of decomposition of the Gini coefficient to solve the problem – the overlap of high rural income residences and low urban income residences. This method decomposed the income inequality of the whole country into four parts: rural areas, urban areas, rural subgroups, and urban subgroups. However, this method does not consider the importance of income disparity between the urban and rural areas. In summary, the question of how to accurately calculate the G_0 is the key element in estimating the Chinese Gini coefficient.

Some other studies calculated the Gini index from other perspectives. Most studies showed that Chinese income inequality from 1983 to 1984 was the lowest since 1978, too. After weighted combination of the data from Renwei Zhao and Shi Li (1997), Dan Huang and Shumin Xi (1999) concluded that the Chinese Gini ratio from 1978 to 1984 had increased. What is more, their Gini coefficient of the whole country is much lower than that of the major studies. Some were even lower than that of the intra-rural Gini ratio conducted by the Chinese Rural Socio-economic Household Surveys Team. Zuguang Hu (2004) used a simple method that has certain demands for the data resources. This method can only provide a very general estimation. As far as formula (4) has been concerned, there is no effective way to calculate G_0 after reviewing current studies on the decomposition of the Gini coefficient.

3. Influence of G_0 on the nationwide Gini coefficient

The early studies on the influence on G_0 were not clear and it was usually called the residual or income overlap part. Bhattacharya and Mahalanobis (1967) called it the concentrated area of income. Mookherjee and Shorrocks (1982) called it a tricky interactive effecting equation that was impossible to calculate accurately. Lambert and Aronson (1993) believed that the residual was the result of the common effects both among groups and between groups, which presented the overlap of income distribution in different groups. Cowell (2000) also put forward similar views. The studies mentioned above proposed that the decomposition of the Gini ratio was influenced by G_0. Li Hu (2005) pointed out that the overlap part among different groups will be small if there are huge income gaps between different groups. The mathematical expression of G_0 from Lambert and Decoster (2005) was:

$$G_0 = 2P_u P_r \frac{\int [1-F_u(x)]F_r(x)dx}{u} \quad (6)$$

Here $F_u(x)$ and $F_r(x)$ are the income distribution function of urban residents and rural residents, respectively, and μ is the average income of residents all over the country. If the population of city and country is equal, G_0 accounts for 50 percent of the overall Gini ratio. If all the Gini coefficients are decomposed according to rural and city residents, and their income distribution follows a certain statistical distribution, the overlapping degree of income of urban and rural residents is closely related to the proportion of income of urban and rural residents and the ratio of their population. Because the current data sources in China limit related research, the calculation of overlapping income between rural and city residents (G_0) has not been found in the relevant literatures.

(1) Distribution function of income

In order to study the influence of G_0 on the calculation of the nationwide Gini coefficients, the distribution of China's urban and rural residents' income will be fitted firstly. However, before fitting residents' income distribution, we need to know what type of distribution function can be used to depict residents' income distribution. Steyn (1950) claimed that the income distribution of the same characteristic people, such as rural residents or urban residents, could be described soundly by lognormal distribution. Balintfy and Goodman (1973) emphasised that income distribution was generated by a special random process which can be explained by using lognormal distribution. Bangwen Cheng (2005) pointed out the universality of the lognormal distribution among the distribution of social-economic scale indicators, and argued that, under the condition of lognormal distribution, the Gini coefficient is determined only by standard deviation. This can also be proved with some empirical results. The World Bank, in its study launched in 2006, proved that income distribution was the lognormal distribution with the data of residents' income in both developed and developing countries over nearly 40 years (Lopez and Servén, 2006). In the study of Japanese residents' income from 1887 to 1998, Souma (2000) pointed out that lognormal distribution is the universal structure of resident income distribution. Holzmann et al. (2007) also utilised lognormal distribution to fit income distribution in their analysis of global income inequality and poverty during the period 1970-2003 according to different development levels of economies and different regions. Xingjian Hong and Jinchang Li (2006) presented a different view, that income distribution could be classified generally into 'pyramid' type, 'dumbbell' type and 'olive' type, which may not all follow lognormal distribution. In reality, however, in China the pyramid and olive types can hardly be found. If the income distribution is an 'olive type', for example, its mode should be close to its mean, which does not correspond to China's resident income distribution. Due to the dual social structure in China, there exists a great gap between urban and rural income. The income distribution of all residents is in fact more similar to a 'dumbbell type'.

However, there are few empirical studies of Chinese resident income distribution function in related literature. The main problem is that at present there is no open raw household survey data. *China Statistics Yearbook* only provides very limited grouping data. The distribution function based on the data has little reliability. In line with our literature review, we found that there were several functions which describe resident income distribution: lognormal distribution, exponential distribution, gamma distribution and Dagum distribution. On the basis of urban household survey data provided by Sichuan Statistical Bureau in 2008, which includes 10,925 household samples, this research studies resident income distribution function.
The distribution of random variables can be described by their PDF (probability density function) or CDF (cumulative distribution function). For the above four distributions, their PDF can be indicated as follows:

$$p(x)=\begin{cases}\frac{1}{\lambda}e^{-x/\lambda}\\ \frac{1}{\sqrt{2\pi}\delta x}\exp(-\frac{\ln^2(x/\mu)}{2\delta^2})\\ \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\\ \beta\lambda\delta t^{-\delta-1}(1+\lambda t^{-\delta})^{-\beta-1}\end{cases} \quad (7)$$

Among the four CDFs, the CDF of exponential distribution is monotonic decreasing, the CDF of lognormal distribution is a little to the right, and the distribution configuration of gamma distribution and Dagum distribution will change significantly as the parameters change.

Firstly, we compared the above four distributions based on the urban resident disposable income data in 2008, which were provided by Sichuan Statistics Bureau. Due to the inconsistent situation between actual population proportions of different areas and the samples' proportions of different areas, we determined the weight of every sample according to the real population in all urban areas. Because of the existence of weight, direct K-S test is not suitable. If we revert the data according to the weight of sample, as the weight should be accurate to the fourth decimal place, this would make the sample number enlarge 10,000 times. As the sample number was so huge, the software could not deal with the samples, so we used another method to deal with the data.

Through comparing the distribution function based on empirical data with the one based on theory, the main purpose of our study is to find the function which has the best goodness-of-fit. To take exponential distribution as an example, for any fractile x and a certain parameter λ, we can write out the theoretical distribution function of such a fractile:

$$F_t(x)=\Pr(X\le x)=\int_0^x \frac{1}{\lambda}e^{-\frac{1}{\lambda}t}dt \quad (8)$$

Then, we arrange the urban resident research data of Sichuan province in 2008 from low to high, and the order of sample weight has changed accordingly. With this treatment, for any x, we assume that n is the maximum index of the income below x:

$$y(n)\le x, y(n+1)>x \quad (9)$$

Therefore, the distribution function of empirical data is:

$$F_i(x)=\frac{\sum_{k=1}^{n}w_k}{\sum_{k=1}^{M}w_k} \quad (10)$$

The criterion which we use to measure the accuracy of goodness-of-fit is the total error between theoretical distribution function and empirical distribution function.

$$\sigma = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[F_t(x_k) - F_i(x_k)]^2} \quad (11)$$

Through the adjustment of λ, we can find a λ to make the error minimum, then we take this λ as our estimate parameter. In the same way, we can estimate the parameters and errors of the other three distributions. The results are as follows:

Table 1. Main parameters and goodness-of-fit error of kinds of distribution function

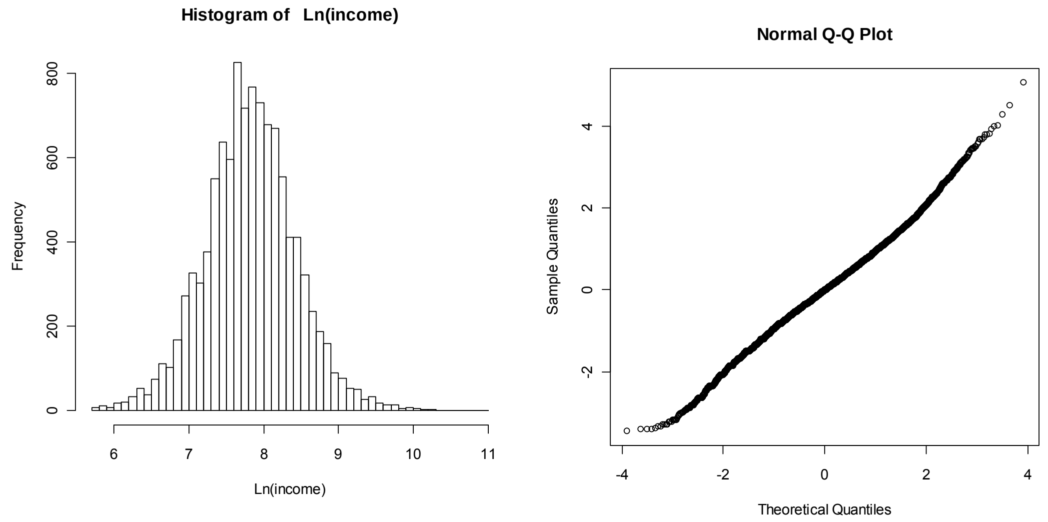
Function name	Exponential distribution	Lognormal distribution	Gamma distribution	Dagum distribution
Parameter	$\lambda = 3049.09$	$\mu = 7.8173$ $\delta = 0.5977$	$\alpha = 2.9$ $\beta = 989.05$	$\beta = 250.65$ $\lambda = 86660.29$ $\delta = 2.2191$
Minimum error σ	0□039	0□0039	0□009	0□0093

Table 1 indicates that comparatively lognormal distribution fits the raw household survey data better. Although there is one more parameter in Dagum distribution, its goodness-of-fit is less than lognormal distribution.

To approve this result, we adapt natural logarithm to income, then make out the histogram and the normal Q-Q plot. The results are shown in Figure 1 below. Chart 1 (left) is the histogram which adapts logarithm of the original data. From the chart, we found that it is very similar to the histogram of normal distribution. Furthermore, we compared its real quantile with the theoretical quantile of normal distribution. In the situation of identification, the chart should be a line which came through the origin and the slope should be 1. Chart 2 (right) shows that the data we counted were almost identical to theoretical lognormal distribution. Additionally, we accessed the urban and rural household survey samples of Chengdu in 2008. When all household survey samples were mixed, the samples did not obey lognormal distribution, but looked like a 'dumbbell type'. However, if we study urban and rural resident income distribution separately, we find that they all obey lognormal distribution.

As mentioned above, according to the *Statistics Yearbook*, only grouped household survey data can be found in current published yearbooks. The grouped data are over-concentrated, which brings great difficulty in precisely predicating the Gini ratio. To solve the problem, we have to generate household income data based on current data sources, which will replace the raw household survey data. It will be not difficult to generate household income data if variance and expectation of raw household survey

Figure 1. Histogram and normal Q-Q plot of income distribution



data are known. Here, the EM algorithm is employed to estimate two parameters: variance and expectation.

Expectation-maximisation (EM) algorithm, as an iterative method (Ming Deng and Yi Yang, 2004), is mainly utilised to find the mode of posterior distribution, i.e., maximum likelihood estimates. Iteration consists of two steps: the E-step (expectation) and the M-step (maximisation). The basic idea of EM algorithm is this: with θ as the unknown parameter, we can use Matlab to iterate the two steps until □θ^{i+1}-θ□is minimised. The advantages of EM algorithm are simple and stable. Let us suppose the i^{th} resident's income x_i (i = 12, □, n) is an independent and identically distributed random variable, and obeys lognormal distribution with μ and δ as parameters. The density function is:

$$p(x;\mu,\delta)=\frac{1}{\sqrt{2\pi}\sigma x}\exp(-\frac{\ln^2(x/\mu)}{2\sigma^2}) \quad (12)$$

With the iterative algorithm, μ and δ can be estimated. Then, we can use Matlab to produce the random numbers that obey lognormal distribution with parameters μ and δ. These random numbers represent each individual resident's income, which can approximately fit the real data on urban resident income. Following equation (1), the Gini coefficient can be calculated. In the same way, the density function of rural resident income in the same period and the rural Gini coefficient can be obtained.

Controlling the amount of random numbers on the basis of the real urban and rural population, and merging the urban and rural income distributions to get the national resident income distribution, can help calculate the national Gini coefficient, the intra-rural Gini coefficient, the intra-urban Gini coefficient and the urban-rural Gini coefficient. G_0 can be obtained either by using equation (6)[9] or through the difference between

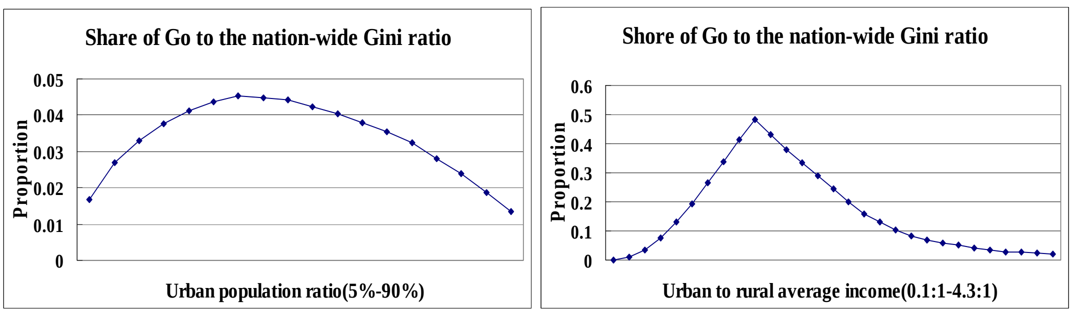
[9] As $F_u(x)$ and $F_r(x)$ are so complex, it is different to calculate G_0 by equation (6).

equation (4) and equation (3). Using the method, we analysed the statistical data of 2005, which showed that the Gini ratio of urban and rural resident income is relatively close to the intra-urban Gini ratio or the intra-rural Gini ratio estimated by the Urban Socio-Economic Survey Corps and the Department of Rural and Social Economic Survey of National Bureau of Statistics of China. Our focus here is upon the exploration of the change trend of G_0 with these findings, instead of an accurate prediction of the national Gini index. Moreover, it is worth noting that the 20 grouped data on the urban resident income can only be traced back to 2005. We cannot, therefore, fit the distribution of urban resident income before 2005. Based on the previous results and the real data in 2005, and with the control of some other variables, we have observed how the change of proportion of urban population and the change of the urban-rural per capita income ratio would influence G_0. The following results are all acquired through Matlab programming.

(2) Main features of G_0

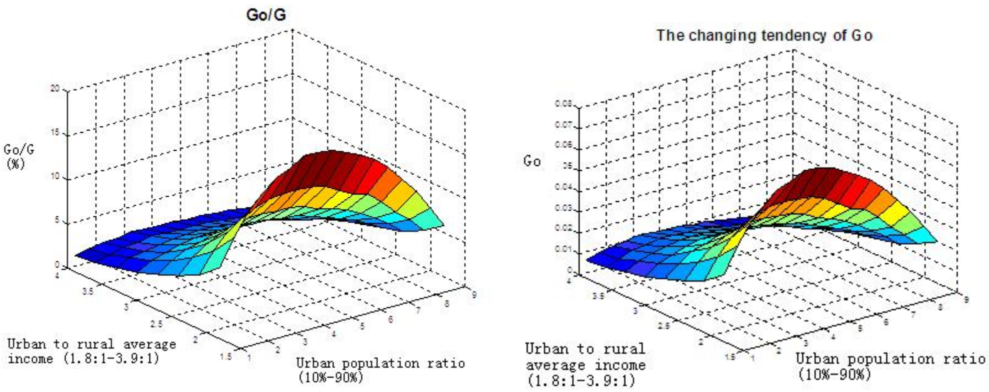
Suppose the share of urban population rises from five percent to 90 percent, Figure 2 (left) demonstrates the 'first-rising-then-descending' trend of the share of G_0 in the national Gini coefficient. As the urban population rises from five percent to 90 percent, G_0 rises from 0.0071 to 0.0214, and then drops to 0.0049; the corresponding proportions in the national Gini coefficient are, respectively, 1.68 percent, 4.54 percent and 1.35 percent, with the highest 4.54 percent corresponding to 30 percent of the urban population.

Figure 2. Main features of *G_0*



Then we can examine how the change of urban-rural resident income affects G_0. Figure 2 (right) demonstrates that as the average income ratio changes from 0.1:1 to 4.3:1, the change trend of G_0 is still firstly rising and then descending, which, in this instance, takes a different form. When the urban-rural average income ratio is 1:1, G_0 is maximised – it makes up nearly half of the Gini coefficient of all residents' income. But as urban-rural resident income increases, the proportion of G_0 in the national Gini coefficient is increasingly diminishing. As the ratio of urban-rural average income is greater than 2.5:1, the proportion of G_0 is less than ten percent. How does G_0 change if the urban population and the urban-rural average income change simultaneously? As illustrated in Figure 3, let us suppose the proportion of urban population changes from ten percent to 90 percent, and the urban-rural income ratio changes from 1.8:1 to 3.9:1:

Figure 3. G_0 trend



As the 3-D graphs (Figure 3) depict, when the proportion of the urban population is around 40 percent and the urban-rural income ratio is 1.8:1, G_0 is maximised; when the former is around 90 percent and the latter 3.9:1, G_0 is minimised. Therefore, the overall trend of the G_0 change is as follows: with an increase in the urban-rural per capita income ratio (R), G_0 descends; with an increase in the proportion of urban population, G_0 first rises and then descends. Moreover, we can analyse the contribution rate of G_0 to all residents' income. Figure 3 (right) demonstrates that when the urban population makes up around 40 percent and the urban-rural per capita income ratio is 1.8:1, G_0 is maximised and, meanwhile, the national Gini coefficient is very low. As a result, the contribution rate of G_0 to the nationwide Gini ratio is up to 18.19 percent. By contrast, as the urban population makes up 90 percent and the urban-rural per capita income ratio is 3.9:1, the contribution rate of G_0 to the nationwide Gini ratio is merely 0.75 percent. According to *China Statistical Yearbook 2006*, the proportion of China's urban population in 2005 was 43 percent; the urban-rural average income ratio was 3.2237:1; and the corresponding G_0 was 0.0189, occupying 4.1 percent of the national Gini coefficient. Assuming all other variables remain unchanging, with an increase in the proportion of the urban population, the G_0/G will descend.

It should be noted that although the proportion of the urban population, and the urban-
distribution variance of the urban and rural resident income is fixed. Therefore, to calculate the national Gini coefficient since Chinese economic reform, it is necessary to measure the variance and the expected value of the distribution of urban and rural

resident income each year. Only in this way can we make an accurate calculation for G_0. Below we introduce methods to measure the variance and the expected value of income distribution of urban and rural residents, and then to calculate the corresponding G_0, and finally to estimate the national Gini coefficient.

4. Estimation of the national Gini coefficient

(1) Method

We present here the main idea of calculating the Gini coefficients of residents' income in China: first of all, with equation (3), the national Gini coefficient with G_0 excluded is calculated. We use the intra-urban Gini coefficient and the intra-rural Gini coefficient provided by the Urban Socio-Economic Survey Corps and the Department of Rural and Social Economic Survey of National Bureau of Statistics of China. Since the survey they carried out can ensure initial data, the Gini coefficient estimated by them is surely reliable. The Gini coefficient between urban and rural areas (G_{ur}) is easy to calculate and not affected by the intra-group income inequality generated by grouped samples, so we can estimate the national Gini coefficient with G_0 excluded. Second, we introduce a new statistic method to calculate G_0. Finally, use equation (4) to calculate the national Gini coefficient. Here we first illustrate how to apply the statistical method to the calculation of G_0.

Suppose μ and δ are, respectively, the mean and the variance of lognormal distribution. Adopting the numerical calculating method, Banwen Cheng (2005) obtained the Gini coefficient corresponding to a different δ. Xingjian Hong and Jinchang Li (2006) proved that, for any income variant x, if Lnx□N (μ□δ^2), then the Gini coefficient:

$$G = 2\Phi(\frac{\delta}{\sqrt{2}}) - 1 \quad (13)$$

With equation (13), for any δ, the Gini coefficient can be easily calculated with the table of standard normal distribution function. On the contrary, here μ and δ are calculated using the intra-urban Gini ratio, the intra-rural Gini ratio and the urban-rural per capita income provided by the National Bureau of Statistics of China. Since $Y = \exp(\mu + \frac{\delta^2}{2})$ (where Y is the known urban-rural average income), equation (14) can be obtained:

$$\begin{cases} \delta = \sqrt{2}\Phi^{-1}(\frac{G+1}{2}) \\ \mu = \ln(Y) - \frac{\delta^2}{2} \end{cases} \quad (14)$$

The intra-urban Gini ratios and the intra-rural Gini ratios from 1978 to 2006, provided by the Urban Socio-Economic Survey Corps and the Department of Rural and Social Economic Survey of National Bureau of Statistics of China, range from 0.15 to 0.3751 (See Table 2), with the range of δ between 0.25 and 1. Through programming with

Matlab, we calculated the corresponding Gini coefficients with δ between 0.25 and 1, and the step length is 0.0001. More than 7,500 Gini coefficients have been calculated, whose range is accurate to the fourth decimal point. Besides, our programming can ensure a corresponding δ after inputting the Gini ratios in urban and rural areas. With δ and the known urban-rural per capita income (Y), μ can be calculated using equation (14). Hence, the distribution function of urban-rural residents' income between 1978 and 2006 can be obtained. Then, with Matlab, we can generate the random numbers that obey lognormal distribution with parameters μ and δ which represent income of an individual resident and can approximately fit the real income data. With the control of random number rates based on real urban and rural population, sample data on national resident income can be obtained. With equation (1), the national Gini coefficient (G) and the national Gini coefficient with G_0 excluded ($G^{'}$) can be easily calculated. Then, since $G_0 = G - G^{'}$, G_0 can be estimated.

All the above calculations are realised by our programming – by inputting the intra-urban Gini coefficient and the intra-rural Gini coefficient of a specific year, the urban-rural per capita income, the proportion of the urban population, the national Gini coefficient and G_0 can be directly outputted. In our real computations, there are in total 10,000 Matlab-generated random numbers that obey lognormal distribution. These numbers are distributed according to the real urban-rural population ratio of that specific year. G_0 is the mean number when a simulation is repeated 100 times. In order to check the convergence of the calculation results, we first get $G^{'}$ for 1978-2006 by using random numbers. Then with equation (3), we get $G^{'}$ for 1978-2006 again, using the intra-urban Gini coefficient, the intra-rural Gini coefficient, the urban-rural par capita income and the urban-rural population ratio provided by the National Bureau of Statistics of China and *China Statistical Yearbook 2006*. We find that $G^{'}$ obtained by each of these two methods is nearly identical, which means that after 100 repetitions of simulation, G_0 is steadily converging to a specific datum.

(2) Advantages of the method

1. The method can eliminate the current restriction of the data source, and can ensure an accurate estimation of China's Gini coefficient. With the initial data from the survey of the residents, the Urban Socio-Economic Survey Corps and the Department of Rural and Social Economic Survey of National Bureau of Statistics of China provide a more accurate estimation on the Gini ratios in urban and rural areas than the *Statistical Yearbook*. Meanwhile, the calculation of G_{ur} is not under the restriction of limited grouped data from the *Statistical Yearbook*. Although the proportion of G_0 in the national Gini ratio is not very high, the current grouped data of the *Statistical Yearbook* cannot accurately reflect the overlapping degree of urban-rural resident income. Using the statistical method we provide here, however, can solve this problem. Therefore, the method can

overcome the current deficiency in data resource and enable us to calculate the Gini coefficient of Chinese residents' income.

2. The method can be used to make a quantitative analysis of the structure of the national Gini coefficient. The decomposed Gini coefficients include the intra-urban Gini coefficient, the intra-rural Gini coefficient and the Gini coefficient between the urban and rural areas. So we can calculate quantitatively the influence of each element on national income inequality, and find the leading factor.

3. Another advantage is that it can be used to make a continuous calculation for national Gini ratios since Chinese economic reform. Because the intra-urban Gini coefficient, the intra-rural Gini coefficient, the urban-rural resident per capita income, and the urban-rural population ratio have already been offered by the National Bureau of Statistics of China and *China Statistical Yearbook 2006*, it is quite convenient to use our method to make a continuous calculation for the national Gini ratios since 1978, and, hence, to make an analysis of the change trend of income inequality since Chinese economic reform.

(3) Results

Table 2 demonstrates that the national Gini ratio of 2006 is 1.52 times more than that of 1978. G_r, G_u and G_{ur} have risen respectively to 76 percent, 110 percent and 56 percent from 1978 to 2006, during which period the contribution rate of G_0 to the Gini ratio of national residents' income is between 0.46 percent and 7.55 percent. Generally, the contribution rate of G_0 is small, but in 1983 and in 1985 the contribution rate of G_0 to the national Gini ratio exceeded the contribution rate of G_0 to the Gini ratio in urban areas. The mid-1980s is exactly the period when the urban-rural income gap became the smallest since Chinese economic reform. As a result, urban and rural income have a large degree of overlap, which leads to a large value of G_0. As discussed above, with the increase of the proportion of the urban population, the contribution rate of G_0 to the Gini coefficient of national residents' income has a trend of 'rising first and then descending'. This trend is in accordance with the change trend of G_0 from the early 1990s to 2006. If we compare G_0 obtained with this method, and G_0 obtained by fitting urban-rural income distribution based on the rural and urban residents grouped data, the difference is only 0.0013. In Table 2, G_r (the intra-rural Gini coefficient) in column (1) and G_u (the intra-urban Gini coefficient) in column (2) are directly from the *China Yearbook of Rural Household Survey* (2001-2007) and *China Statistical Yearbook of Price and Urban Household Survey* (2001-2007) by the Urban Socio-Economic Survey Corps and the Department of Rural and Social Economic Survey of National Bureau of Statistics of China.

Table 2. Gini coefficient and its decomposition, 1978-2006

Year	*Gr*	*Gu*	*Gur*	*G'*	*Go*	*G*	*Go/G* (%)	*βGr/G* (%)	*δGu/G* (%)	*Gur/G* (%)
1978	0.2124	0.16	0.1803	0.3029	0.0014	0.3043	0.46	36.70	3.39	59.25
1979	0.2245	0.16	0.1802	0.3073	0.0021	0.3094	0.68	37.42	3.53	58.24
1980	0.2407	0.16	0.1813	0.3144	0.003	0.3174	0.95	38.19	3.67	57.12
1981	0.2406	0.15	0.1597	0.294	0.0044	0.2984	1.47	41.12	3.66	53.52
1982	0.2317	0.15	0.1355	0.2665	0.007	0.2735	2.56	43.64	4.02	49.54
1983	0.2461	0.15	0.1183	0.257	0.0123	0.2693	4.57	47.67	4.03	43.93
1984	0.2439	0.16	0.1241	0.2587	0.0124	0.2711	4.57	44.73	4.81	45.78
1985	0.3072	0.19	0.1291	0.2927	0.0239	0.3166	7.55	46.92	5.21	40.78
1986	0.3042	0.19	0.1629	0.3182	0.0158	0.334	4.73	40.69	5.69	48.77
1987	0.2889	0.2	0.17	0.3159	0.0142	0.3301	4.30	37.70	6.49	51.50
1988	0.3053	0.23	0.1717	0.3263	0.0196	0.3459	5.67	37.34	7.38	49.64
1989	0.3185	0.23	0.1864	0.341	0.0184	0.3594	5.12	36.06	7.52	51.86
1990	0.3099	0.23	0.1772	0.3319	0.0189	0.3508	5.39	36.32	7.64	50.51
1991	0.3072	0.24	0.1984	0.3491	0.0155	0.3646	4.25	32.66	8.32	54.42
1992	0.3134	0.25	0.2205	0.369	0.0142	0.3832	3.71	29.98	8.86	57.54
1993	0.3292	0.27	0.2412	0.3928	0.0142	0.407	3.49	27.92	9.67	59.26
1994	0.321	0.3	0.2483	0.4004	0.015	0.4154	3.61	25.79	10.98	59.77
1995	0.3415	0.28	0.2358	0.3935	0.0177	0.4112	4.30	27.92	10.41	57.34
1996	0.3229	0.284	0.2172	0.3714	0.0199	0.3913	5.09	27.30	11.60	55.51
1997	0.3285	0.292	0.214	0.3699	0.0225	0.3924	5.73	26.43	12.74	54.54
1998	0.3369	0.3	0.2189	0.3783	0.0243	0.4026	6.04	24.73	13.83	54.37
1999	0.3361	0.295	0.2377	0.3889	0.0203	0.4092	4.96	22.20	14.68	58.09
2000	0.3536	0.319	0.2506	0.4099	0.0222	0.4321	5.14	20.21	16.39	58.00
2001	0.3603	0.323	0.26	0.4173	0.0217	0.439	4.94	18.59	17.64	59.23
2002	0.3646	0.32	0.2754	0.4321	0.0185	0.4506	4.11	16.44	18.50	61.12
2003	0.368	0.34	0.2824	0.4467	0.019	0.4657	4.08	14.68	20.35	60.64
2004	0.3692	0.3245	0.2794	0.4394	0.0173	0.4567	3.79	14.24	20.71	61.18
2005	0.3751	0.3198	0.2786	0.4397	0.0176	0.4573	3.85	13.62	21.31	60.92
2006	0.3737	0.336	0.2805	0.4448	0.0176	0.4624	3.81	12.72	22.95	60.66

Notes:

1. G_{ur} in column (3) is from calculations by the author, based on data in the *China Statistical Yearbook of 2007* provided by the National Bureau of Statistics of China.
2. G' (the national Gini coefficient) in column (4) is from calculations by the author using equation (1).
3. G_0 in column (5) is from calculations by the author using statistical method.
4. G' (the national Gini coefficient) in column (4) is from calculations by the author using equation $G= G'+ G_0$.
5. Columns (7), (8), (9) and (10) are, respectively, the percentage shares of G_0, the intra-rural Gini coefficient, the intra-urban Gini coefficient, and the Gini coefficient between urban and rural areas, in the national Gini coefficient.

The contribution rate of the intra-rural Gini coefficient to the national Gini coefficient decreases from 36.7 percent in 1978 to 12.72 percent in 2006; during the same period, the contribution rate of the intra-urban Gini coefficient to the national Gini coefficient rises from 3.39 percent in 1978 to 22.95 percent in 2006. Following this trend and the urbanisation, the contribution rate of the intra-urban Gini coefficient to the national Gini coefficient will be increasingly enhanced. Although G_r and G_u have risen at a rapid speed, their contributions to the national Gini coefficient are limited. The key factor manipulating the trend of the national Gini coefficient is the Gini coefficient between urban and rural areas (G_{ur}). As Table 2 shows, G_{ur} is always the determining factor of the national Gini coefficient, and in 2006 it even made up 60.66 percent of the national Gini coefficient. In addition, the two bear the same change trend. Therefore, considering the dominating effect of G_{ur} on the change of the national Gini coefficient and the same change trend, we should now shift our attention to the urban-rural income gap.

5. Influence of G_{ur} on the national Gini coefficient

Suppose the urban-rural per capita income ratio is R, the Gini coefficient between urban and rural inhabitants (G_{ur}) can be expressed as (deductions omitted):

$$G_{ur} = \frac{(R-1)P_u(1-P_u)}{(R-1)P_u+1} \quad (15)$$

With the same R, after derivation, the max of G_{ur} (i.e., G_{urmax}) is:

$$G_{ur\max} = \frac{(\sqrt{R}-1)(R-\sqrt{R})}{\sqrt{R}(R-1)} (16)$$

With the increase of the urban population (P_u), the Gini ratio between urban and rural areas becomes larger. When G_{ur} reaches the max (G_{urmax}), with the further increase of P_u, G_{ur} will descend. Therefore the U-curve can be found. For $\partial G_{ur} / \partial R \geq 0$ and $\partial^2 G_{ur} / \partial R^2 < 0$, the curve has a crest in the upper left field, which means that with the increase of R, the influence of R on G_{ur} is dropping. Therefore, urbanisation is crucial to the diminishment of urban-rural income inequality.

The above study demonstrates that the key factor for determining the trend of the Chinese Gini coefficient is the Gini coefficient between urban and rural areas. It is necessary, therefore, to explore how the urbanisation process, and the change of urban-rural per capita income, affect national income inequality.

(1) Influence of urbanisation process

With the data on the intra-urban Gini coefficient and the intra-rural Gini coefficient, which are provided by the Urban Socio-Economic Survey Corps and the Department of Rural and Social Economic Survey of National Bureau of Statistics of China in 2005, and with the data on the urban and rural population proportions and the urban-rural per capita income, which are provided by *China Statistical Yearbook 2006*, according to formula (14), we can calculate the main parameters of the income distribution of urban and rural residents in 2005. With the income distribution function of urban-rural residents, the influence of the urbanisation process and the change of the urban-rural per capita income ratio on national income inequality can be explored. First, without changing the values of μ and δ, this influence can be observed through increasing the proportion of the urban population from five percent to 90 percent.

Figure 4. Influence of the urbanisation process and the urban-rural income gap

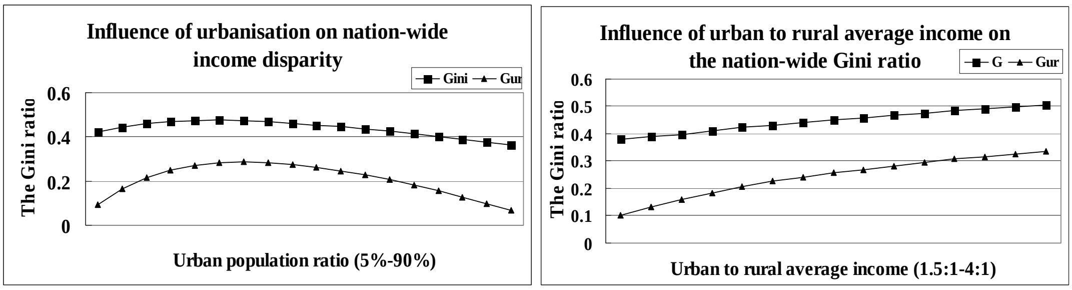


Figure 4 demonstrates the mean number in 100 repeated simulations of the 10,000 computer-generated random numbers with the known distribution function. Figure 4 (left) shows that as the proportion of the urban population rises from five percent to 90 percent, the national Gini coefficient rises from 0.4230 to 0.4755 and then drops to 0.3613, while the Gini coefficient between urban and rural areas rises from 0.0951 to 0.2846 and then drops to 0.0667. The same change trend can be found here: with urbanisation, the income inequality rises first and then descends. The proportion of urban population corresponding to the highest national Gini coefficient is around 26 percent, and the proportion of the real urban population in 2005 reaches 43 percent, which means that further urbanisation will contribute to the national income equality.

In line with Lewis's dual economy theory (Lewis, 1954), during the initial period of industrialisation, urban industrial sectors may gain the most profits from society. As a result, income inequality between urban industrial sectors and rural traditional sectors increases. With more and more rural population transferring to urban areas, there could be no surplus rural labour. Consequently, real rural income and all society's average income will be enhanced, which will surely bring in the diminishment of the urban-rural income inequality. Based on Lewis's dual-sector model, Kuznets (1955), by reviewing the economic development of developed countries, holds that the evolution of income inequality is an invert U pattern, i.e., rising first and then descending. Therefore, urbanisation exerts a critical effect on national income inequality.

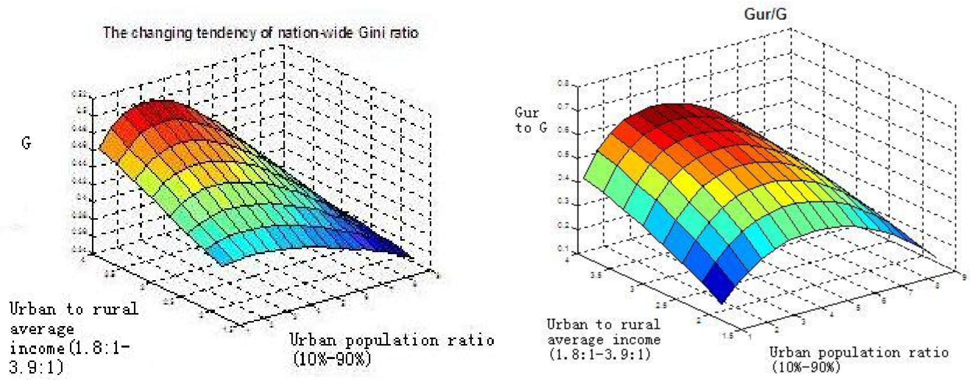
(2) Influence of urban-to-rural per capita income ratio

According to the data for the period from 1978 to 2006 in the *China Statistical Yearbook*, in 1983 the urban-rural per capita income ratio arrived at its lowest, 1.8225:1; in 2007, it reached its highest, 3.33:1. With the 2005 data, and without changing the parameter δ, we change the ratio from 1.5:1 to 4:1 to observe the change in national income inequality. According to Figure 4, when the ratio is changed from 1.5:1 to 4:1, the national Gini coefficient rises from 0.3793 to 0.5051, and the Gini coefficient between urban and rural areas rises from 0.1009 to 0.3211. The influence of the rise of the urban-rural per capita income ratio on the urban-rural Gini coefficient exceeds the influence on the national Gini coefficient; the urban-rural Gini coefficient's share in the national coefficient also rises from 26.6 percent to 63.6 percent. We also discover that the rising of the urban-rural per capita income ratio may increase the income inequality, but this influence has a convergent tendency: as discussed above, for $\partial G_{ur} / \partial R \geq 0$ and $\partial^2 G_{ur} / \partial R^2 < 0$ with the value of *R* becoming larger, the growth rate of the Gini coefficient between urban and rural areas becomes smaller. If expanding indicators of urban-rural per capita income, the following can be clearly observed: G_{ur} is a curve with a crest in the upper left field, i.e., with a continuous increase in urban-rural per capita income ratio, the growth rate of the urban-rural Gini coefficient has a downward tendency.

(3) Co-influence of population and urban-to-rural per capita income ratio

If the proportion of urban population and the urban-rural per capita income ratio change simultaneously, how does the national Gini ratio change? Here, we change the urban population share from ten percent to 90 percent and the urban-rural per capita income ratio from 1.8:1 to 3.9:1. Figure 4 (left) demonstrates that, with the increase in the proportion of urban population, the national Gini ratio first rises and then descends, and with the rise in the urban-rural per capita income ratio, the proportion of urban population corresponding to the highest value of the Gini ratio of national resident income also increases. Take the year 2005 as an example. When the proportion of urban population reaches 26 percent, a further increase in urban population will contribute to the diminishment of national income inequality: the national Gini ratio will peak only when the urban-rural income ratio rises to 3.9:1 and the urban population occupies 30 percent. According to the *Statistical Communiqué of the People's Republic of China on the 2008 National Economy and Social Development*, the proportion of urban population in 2008 is 45.68 percent, and the urban-rural per capita income ratio is 3.3146:1. Obviously, further urbanisation contributes to the improvement of national income distribution.

Figure 5. National Gini index and G_{ur}/G trends



Due to the key role played by the urban-rural income gap in the national income gap, we here explore the change in the share of urban-rural income inequality in national income inequality. Figure 5 demonstrates that, with an increase of the urban-rural income ratio, the share of G_{ur} in the national Gini coefficient increases too. Meanwhile, with the enhancement of the proportion of urban population, the share of G_{ur} in the national Gini coefficient first rises and then descends. In this instance, when urban population constitutes 35 percent and the urban-to-rural per capita income ratio is 3.9:1, G_{ur} takes the largest share of 65.39 percent of the national Gini coefficient; by contrast, when urban population constitutes 90 percent and urban-rural per capita income ratio is 1.8, G_{ur} takes the smallest share of the national Gini coefficient – only 12.18 percent. Even if the urban-rural per capita income ratio remains 3.9:1 and the proportion of the urban population is as high as 90 percent, G_{ur} takes only 19.75 percent of the national Gini coefficient. Therefore, urbanisation is beneficial to the decrease of the share of G_{ur} in the national Gini coefficient.

6. Conclusions and policy suggestions

The data source at present is responsible for the puzzles concerning China's income Gini index. In order to overcome the data source limitation, much current research attempts to make a breakthrough by improving calculating methods, especially through decomposing the national Gini ratio by urban and rural areas. However, there exists an overlap between urban and rural resident income, which can affect the accuracy of the calculation of the national Gini coefficient. This paper studies the characteristics and influences of G_0 by applying the two-step EM algorithm to fit China's urban and rural resident income distribution in 2005. Based on that, we compiled a program to analyse how the change of population and urban-rural per capita income ratio affect the value of G_0. Moreover, given the present condition of the data source, we calculated the Chinese Gini coefficient since China's economic reform using the statistical method. Through the decomposition of the national Gini coefficient, we find that the Gini coefficient between urban and rural residents is the key factor in affecting China's income inequality, and that urbanisation and the change of urban-rural per capita income ratio are the key factors affecting urban-rural income inequality. As a result, the paper made a deeper exploration of the influence of the change of urban population and urban-rural per capita income ratio on national income inequality. It turns out that urbanisation will cause the national Gini coefficient and the Gini coefficient between urban and rural areas to rise first and then descend; and although an increase in the urban-rural per capita income ratio will enhance the degree of income inequality, its influence takes a convergent trend. Therefore, diminishing urban-rural income inequality is the key factor in improving China's income equality, and speeding up urbanisation is the key factor in reducing urban-rural income inequality.

There are two ways to reduce income equality between urban and rural areas. The first is to increase farmers' income and thus reduce the urban-rural income gap. For instance, China has implemented a series of reforms, such as abolishing agricultural tax, direct grain subsidy, seed subsidy, agricultural machinery subsidy, rural land circulation and the like. The second way is to enhance urbanisation. According to the National Bureau of Statistics of China, in 2007, the contribution rate of primary industry to national GDP was merely 11.26 percent, whereas the share of the labour force in primary industry occupied 40.8 percent of national labour resources. If, in 20 or 30 years' time, there are still such large numbers of farmers relying on limited land resources, rural income cannot be increased significantly. Therefore, it is very hard to break the urban-rural dual social structure if only the question of how to increase rural income is considered. The significance and urgency of breaking the urban-rural dual social structure has already been manifested by the judgements of the third Plenary Session of the 17th Communist Party of China Central Committee, that China has arrived at an important period for breaking the urban-rural dual social structure and bringing in a new era of uniform economic and social development for urban and rural areas. In other words, the key to breaking up the dual structure is to promote the urbanisation process in China.

As far as the present situation is concerned, the government should propel the construction of western city clusters, especially, the Chengdu-Chongqing city cluster. The western district holds surplus rural labour that can be absorbed into the construction of city clusters. The employment pressure from the large-scale transfer of rural labour and the infrastructure pressure can be relieved by the building of city clusters and scattered small and medium-sized towns. With the opportunities for industrial upgrading of the eastern district, investment from the eastern industries can be introduced. In terms of industrial development, labour-intensive industry and especially the service industry should be supported, since there will be some difficulties and obstacles in transferring education- and skill-limited rural labour directly into the modern sectors in the cities; by contrast, with a low entrance threshold, the service industry can absorb a large number of rural labourers.

There is, furthermore, another effective way to improve urban-rural income inequality, namely, to transfer rural labour from the remote less-developed areas in the middle and western parts of China to medium-small-sized cities or economically developed areas. This strategy will enhance not only the urbanisation process but also the average income of rural residents. With the decline of rural poor population, the rural average income will be raised and the urban-rural per capita income gap will be diminished. Admittedly, the smooth transfer of trural surplus labour relies on the capacity to create more employment opportunities that obviously derive from economic development. Surely economic growth cannot automatically solve China's income inequality, but income inequality is difficult to improve without it.

References

Balintfy, L. and Goodman, S. (1973). 'Socio-economic factors in income inequality: A log-normal hypothesis'. *Zeitschrift fiir Nationalokonomie*, 33, 389-402.

Bhattacharya, N. and Mahalanobis, B. (1967). 'Regional disparities in household consumption in India'. *Journal of the American Statistical Association*, 62, 143-161.

Champernowne, D. and Cowell, F. (1998). *Economic Inequality and Income Distribution*. Cambridge, UK: Cambridge University Press.

Chen, Jiandong, Jin, Shengwu and Tang, Xinyue (2005). 'A new method of analyzing urban-rural income gap and an analysis on 1986-2003 China's urban-rural inequality'. In Cai, Fang and Wan, Guanghua (eds.), *Studies on Income Inequality and Poverty in Transitional China*. Beijing: Social Sciences Academic Press.

Chen, J., Hou, W. and Jin, S. (2008). 'The Chinese Gini coefficient and its decomposition'. Paper presented at the 16th Annual Conference on PBFEAM, Brisbane, Australia, 2-4 July.

Chen, Zongsheng (1999). *Reform, Development and Income Distribution.* Shanghai: Fudan University Press.

Chen, Zongsheng and Zhou, Yunbo (2002). *Rethinking on Income Distribution in Reform and Developemnt: China Has a Bipolar Economy?* Beijing: Economic Science Press.

Cheng, Bangwen (2005). 'Lorenz curve and Gini coefficient under condition of lognormal distribution'. *Journal of Quantitative and Technical Economics*, 2, 127-135.

Cheng, Yonghong (2007). 'Overall Gini coefficient of China and its decomposition by rural and urban areas since reform and opening-up'. *Social Sciences in China*, 4, 45-60.

Cheng, Yonghong (2006). 'Calculation and decomposition of the overall Gini coefficient in dual economies'. *Economic Research Journal*, 1, 109-120.

Cowell, F. (2000). 'Measuring inequality'. In A. Atkinson and F. Bourguignon (eds.), *Handbook: Income Distribution*. Amsterdam, New York and Oxford: Elsevier North-holland.

Deaton, A. (1995). 'Data and econometric tools for development analysis'. In J. Behrman and N. Srinivasin (eds.), *Handbook of Development Economics*, 3. Amsterdam: Elsevier Press.

Deng, Ming and Yang, Yi (2004). 'Estimating parameter in lognormal distribution from grouped data'. *Systems Engineering –Theory Methodology Application*, 6, 553-556.

Department of Rural and Social Economic Survey of National Bureau of Statistics of China (2000-2007). China *Yearbook of Rural Household* Survey. Beijing: China Statistics *Press.*

Fang, C., Zhang, X. and Fan, S. (2002). 'Emergence of urban poverty and inequality in China: Evidence from household survey'. *China Economic Review,* 13, 430-443.

Gibson, J., Huang, J. and Rozelle, S. (2001). 'Why is income inequality so low in China compared to other countries? The effect of household survey methods'. *Economics Letters*, 71, 329-333.

Holzmann, H., Vollmer, S. and Weisbrod, J. (2007). 'Income distribution dynamics and pro-poor growth in the world from 1970 to 2003'. *Ibero America Institute for Economic Research (IAI) Discussion Papers*, 161.

Hong, Xingjian (2008). 'A new subgroup decomposition of the Gini coefficient'. *China Economic Quarterly*, 1(8), 307-342.

Hong, Xingjian and Li, Jinchang (2006). 'A further study about some problems of the Gini coefficient', *Journal of Quantitative and Technical Economics*, 2, 86-96.

Huang, Dan and Xi, Youmin (1999). 'Digression atmosphere of marginal propensity to consume'. *Economic Research Journal*, 5, 16-19.

Khan, A. and Riskin, C. (2001). *Inequality and Poverty in China in the Age of Globalisation.* Oxford, UK: Oxford University Press.

Kuznets, S. (1955). 'Economic growth and income inequality'. *American Economic Review*, 45, 1-28.

Lambert, P. and Aronson, R. (1993). 'Inequality decomposition analysis and the Gini coefficient revisited'. *The Economic Journal*, 103, 1221-1227.

Lambert, P. and Decoster, A. (2005). 'The Gini coefficient reveals more'. *International Journal of Statistics,* LXIII, 373-400.

Lewis, W. A. (1954). 'Economic development with unlimited supplies of labor'. *Journal of the Manchester School of Economics and Social Studies*, 20, 139-192.

Li, Hu (2005). 'A research on Gini coefficient decomposition analysis'. *Quantitative & Technica Economics*, 3, 127-135.

Li, Shi (2003). 'Review and prospect: Studies on China's resident income distribution', *China Economic Quarterly*, 2(2), 379-404.

Li, Shi (2002). ‘A further discussion on estimation and decomposition of Gini coefficient’. *Economic Research Journal*, 5, 84-87.

Li, Shi and Luo, Chuliang (2007). ‘Re-estimating the income gap between urban and rural households in China’. *Journal of Peking University (Philosophy and Social Sciences)*, 44 (2), 111-120.

Li, Shi, Sicular, T, and Gustafson, B. (2008). *Research on Income Dstribution in China.* Beijing: Beijing Normal University Press.

Lin, Yifu, Cai, Fang and Li, Zhou (1998). ‘An analysis on regional gap in China economic restructuring’. *Economic Research Journal*, 10, 3-10.

Lopez, J. and Servén, L. (2006). *A Normal Relationship? Poverty, Growth and Inequality.* Washington, DC: World Bank Policy Research Working Paper 3814.

Milanovic, B. (2002). ‘True world income distribution, 1988 and 1993: First calculation based on household surveys alone’. *Economic Journal,* 112, 51-92.

Mookherjee, D. and Shorrocks, A. (1982). ‘A decomposition analysis of the trend in UK income inequality’. *The Economic Journal*, 92, 886-902.

National Bureau of Statistics of China (1987-2007). *China Statistic Yearbook.* Beijing: China Statistics *Press.*

National Bureau of Statistics of China (1978-2008). *Statistical Communiqué of The People's* Republic *of China on the National Economic and Social Development*. Available online: http://www.stats.gov.cn/tjgb/index.htm (accessed 7 January 2010).

Pyatt, G. (1976). ‘On the interpretation and disaggregation of Gini coefficients’. *Economic Journal,* 86, 243-255.

Ravallion, M. and Chen, S. (2007). ‘China’s (uneven) progress against poverty’. *Journal of Development Economics,* 82, 1–42.

Research Group of Department of *Rural* and Social Economic Survey of National Bureau of Statistics of China (1994). ‘Study on income gap between urban and rural residents in China’. *Economic Research Journal*, 12, 34-45.

The Urban Socio-Economic *Survey Corps* of National Bureau of Statistics of China (2000-2005). China *Statistical Yearbook of Price and Urban Household* Survey. Beijing**:** China Statistics *Press.*

Sen, A. (1997). *On Economic Inequality.* Oxford, UK: Clarendon Press.

Shorrocks, F. (1984). ‘Inequality decomposition by population subgroups’. *Econometrica,* 52, 1369-1385.

Shorrocks, F. and Wan, G. (2005). ‘Spatial decomposition of inequality’*. Journal of Economic Geography,* 5, 59-82.

Sloman, J. (2000). *Economics.* 4th edition. London: Prentice Hall/ Financial Times.

Souma, W. (2000). *Universal Structure of the Personal Income Distribution*. Available online at: http://arxiv.org/PS_cache/cond-mat/pdf/0011/0011373v1.pdf (accessed 8 January 2010).

Steyn, S. (1950). ‘A model for the distribution of incomes’. *South African Journal of Economics*, 27(2), 149-156.

Sundrum, M. (1990). *Income Distribution in Less Developed Countries.* London: Routledge.

The Urban Socio-Economic *Survey Corps* of National Bureau of Statistics of China (2006-2007). *China Urban Life and Price* Yearbook*.* Beijing: China Statistics *Press.*

Wan, Guanghua (2004). ‘Measure and decomposition of income distribution: A review on research methodology’. *World Economic Papers*, 1, 64-69.

Wang, Chunlei and Huang, Suxin (2007). ‘Gini coefficient and sample information content’. *Quantitative and Technical Economics*, 2, 136-144.

Wei, Houkai. (1996) ‘The income gap of residents in different regions in China and its decomposition’. *Economic Research Journal*, 11.

Xiang, Shujian (1998). ‘Calculation and regression analysis on Gini coefficient of Chinese residents’ income’. *The Theory and Practice of Finance and Economics*, 1, 75-80.

Yang, D. (1999). ‘Urban-biased policies and rising income inequality in China’. *The American Economic Review*, Papers and Proceedings of the One Hundred Eleventh Annual Meeting of the American Economic Association, 89(2), 306-310.

Yao, S. (1997) ‘Decomposition of Gini coefficients by income factor: A new approach and application’. *Journal of Applied Economics*, 4, 27-31.

Zhao, Renwei and Li, Shi (1997). ‘The reasons of enlarged income disparity in China’. *Economic Research Journal*, 9, 19-28.

Zhao, Renwei, Li, Shi, and Riskin, C. (1999). *A Restudy on Income Distribution of Chinese Residents*. Beijing: China Financial and Economic Publishing House.

Zuguang Hu (2004). 'A study of the best theoretical value of Gini coefficient and iIts concise calculation formula'. *Economic Research Journal*, 9, 60-69.

The University of Manchester
Brooks World
Poverty Institute

MANCHESTER
1824

Executive Director
Professor David Hulme

Research Director
Professor Armando Barrientos

Contact:

Brooks World Poverty Institute
The University of Manchester
Humanities Bridgeford Street Building
Oxford Road
Manchester
M13 9PL
United Kingdom

Email: bwpi@manchester.ac.uk

www.manchester.ac.uk/bwpi

The Brooks World Poverty Institute (BWPI) creates and shares knowledge to help end global poverty.

BWPI is multidisciplinary, researching poverty in both the rich and poor worlds.

Our aim is to better understand why people are poor, what keeps them trapped in poverty and how they can be helped - drawing upon the very best international practice in research and policy making.

The Brooks World Poverty Institute is chaired by Nobel Laureate, Professor Joseph E. Stiglitz.

www.manchester.ac.uk/bwpi