The trend of the Gini coefficient of China (1978–2010)

Citation for published version:
Chen, J, Pu, M & Hou, W 2019, 'The trend of the Gini coefficient of China (1978–2010)', *Journal of Chinese Economic and Business Studies*. https://doi.org/10.1080/14765284.2019.1663695

Digital Object Identifier (DOI):
10.1080/14765284.2019.1663695

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Chinese Economic and Business Studies

The trend of the Gini coefficient of China (1978–2010)

Abstract

This study proposes a new approach to analyse the effects of an overlap term on the calculation of the overall Gini coefficient and estimates China's Gini ratios since the adoption of the economic reform and open-door policies. A decomposition of the Chinese Gini coefficient for 1978–2010 reveals that the key factor contributing to income inequalities is the income disparity between rural and urban inhabitants. We further investigate the features of this income inequality between rural and urban areas and employ statistical approaches to evaluate the effects of urbanisation and rural-to-urban average income on nationwide income inequality. The results show that accelerating the pace of urbanisation is mainly responsible for decreasing China's income disparity. Drawing on these results, we conclude with suggestions for related policies.

Keywords: Chinese Gini coefficient; Decomposition of Gini coefficient; Urbanisation; Income disparity; Rural–urban income

1. Introduction

Among the various methods used to measure income inequality, such as the Theil index, coefficient of variation, and Kuznets index, the Lorenz curve and Gini coefficient are the most commonly used (Sloman, 2000). Of the two, the Gini coefficient is considered an important index to estimate income inequality (Sen, 1997; Champernowne and Cowell, 1998). In 1912, Corrado Gini, an Italian statistician, published *Variability and Mutability*, in which he proposed a method to measure inequality which gradually evolved into the well-known Gini coefficient (Li, 2002).

However, in the context of China, there are several disagreements regarding the calculation of the Gini coefficient for its residents. In fact, our literature review revealed about 20 different estimations of Chinese Gini ratios. A noteworthy calculation of the Gini ratio for China, nonetheless, is Chen's (1999) estimation of 0.365 for the year 1995. In 2002, however, he and Zhou used two other methods and, respectively, presented values of 0.38392 and 0.41914. Cheng (2006, 2007) and Chen, Hou, and Jin (2008) estimated the value at 0.4169, and 0.3934, while Xiang (1998) and Huang and Xi (1999) obtained values of 0.3515 and 0.328. Zhao, Li, and Riskin (1999) stated it is 0.445. The highest value, 0.452, was provided by Khan and Riskin (2001), which is 9% higher than the 0.415 obtained by Ravallion and Chen (2007) and almost twice the lowest value. Furthermore, the Urban Social Economic Survey Corps (USESC) and the Department of Rural Social Economic Survey (DRSES) of the National Bureau of Statistics (NBS) of China have estimated the intra-rural and intra-urban Gini coefficient since 1978; their results are widely quoted in the literature. NBS has also announced the nationwide Gini coefficient for 2013, which can be traced to 2003. However, there remain disagreements regarding the calculation of the nationwide Gini coefficient, particularly since the adoption of the economic reform and open-door policies. The Gini coefficient provides basic statistical data to analyse income inequality in China. Therefore, such a variation limits the progress of studies on China's income inequality. Further, current studies on China's income disparity tend to examine numerous variables affecting income distribution, but pay less attention to investigating the trend of the Gini ratio for China.

In line with a previous study (Chen et al., 2010), this paper estimates Chinese Gini ratio from 1978 to 2010 and presents a clear picture of the changing tendency of nationwide Gini ratio. The marginal contributions of this study are as follows. First, we present a new approach to estimate the Chinese Gini ratio for 1978–2010 that allows us, to some extent, to avoid the shortcomings of current data sources. Second, using the results of the above estimation, we decompose nationwide Gini ratios, revealing the Gini ratio trend in China. Third, the paper explores the features of overlap term from the decomposed Gini ratio based on rural and urban residents. Fourth, we identify that the primary factor contributing to China's income inequality is the income disparity between rural and urban areas. Meanwhile, urbanisation is a key factor narrowing the Chinese rural–urban income gap, and therefore we suggest that accelerating the pace of urbanisation will help decrease China's income disparity.

The remainder of this paper is organised as follows. Section 2 analyses the shortcomings in the present calculations of the Chinese Gini coefficient on the basis of a literature review. Section 3 examines the influence of overlap terms on the estimation of the Gini ratio. Section 4 demonstrates the estimation of Chinese Gini ratios and decomposes the Gini ratio by city and countryside. Section 5 discusses the effects of the income gap between cities and the countryside on the nation's income inequality with focus on urbanisation and the urban-to-rural income ratio. Section 6 concludes the paper and offers suggestions for policymakers.

2. Studies on the Chinese Gini coefficient

The ideal statistical source to calculate a Gini coefficient is original household survey data. Using see Equation (1), the Gini ratio can be easily and accurately calculated by employing software such as DAD developed by LAVAL, which estimates the index for extreme income inequalities. Li (2002) also highlighted the use of software that specifically calculates inequality indexes such as INEQ or Stata to estimate the Gini ratio. All calculations in this study are performed using Matlab.

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

where n is population (or family count), u is average income, and y_i and y_j denote the income of i family and j family. However, the Chinese income data currently available are not the original data but those divided into groups as per income level. For example, the *China Statistical Yearbook* divides only urban resident samples, which cover more than 60,000 households, into seven groups. Given the small number of groups, it is difficult to guarantee the accuracy of the Gini ratio. By conducting a literature review, in the following subsections, we argue that the shortcomings in estimating the Chinese Gini ratio mainly stem from the aspects outlined below.

2.1 China Statistical Yearbook and similar data sources

Several statistical yearbooks have been published in China, including the *China Statistical Yearbook*, *Yearbook of Urban Living and Price Index*, *Yearbook of the Rural Household Survey* (The yearbook was first published in 1992; however, it was not until 2000 that the NBS began publishing it every consecutive year.), and *Statistical Yearbook of Chinese Price Index and Urban Household Survey of Income and Expenditure*. The last two sample households in a manner similar to that of the *China Statistical Yearbook*. Although the *Yearbook of Urban Living and Price Index* includes a higher number of income groups than that in the *China Statistical Yearbook,* which categorises all samples into 20 groups as per income level, it was officially published only in 2006.

The *China Statistical Yearbook* is the most crucial data resource used to calculate the Chinese Gini ratio, particularly when no household survey data are available. For example, Chen and Zhou (Chen, 1999; Zhou and Chen, 2002) use data from the yearbook to estimate the Gini coefficient. However, the income provided in the yearbook has been greatly doubted. Khan and Riskin (2001) commented that the statistical data are too aggregated and hinder the careful analysis of income inequality. Fang, Zhang and Fan (2002) also believed that the aggregated data ignored the income disparity within each group and are not sufficiently accurate. In 1996, the *China Statistics Yearbook* stated that those with an income of more than RMB 2,000 constituted 38.4% of all rural households. However, the yearbook did not further divide such households into subgroups. Furthermore, the 2007 yearbook declared that those with an income of more than RMB 5,000 accounted for 30.94% of all rural households. Yet, after almost a decade, these households remain undivided into further subgroups.

Another problem in calculating the Chinese Gini coefficient is that the *China Statistical Yearbook* provides two categorical samples—one from urban areas and the other from rural areas—and fails to integrate them. To resolve this issue, Chen and Zhou (2002) proposed a statistical approach that first adds different weights to rural and urban samples on the basis of actual population ratios and then, combines them. While there is no issue with the method itself, it does not solve the problems identified for the *China Statistical Yearbook*, that is, the over-concentration of groups and the oversight of differences within each group.

Khan and Riskin (2001) asserted that the standard of categorising residents' income in China differs from the international one. According to Li and Luo (Li, 2003; Li and Luo, 2007), the hidden subsidies for urban residents are considerably higher than those for rural ones. For instance, rural residents lack benefits such as housing, medical, pension, and unemployment insurance. Accounting for these hidden benefits increases the income ratio between cities and the countryside by one-third. Therefore, the actual income gap between cities and the countryside is undoubtedly larger than that provided in the yearbook.

In addition, changes in statistical methods have led to obvious inconsistencies and inaccuracies in the data. Income and population are two variables that are essential to calculate the Gini coefficient. However, the yearbook does not provide consistent statistics standards for income and population.

Prior to 1991, the income level was set as average income rather than average disposable income. However, since 1995, NBS has been using average disposable income as its statistics source. Moreover, in the 1994 *China Statistical Yearbook*, the packet number for rural residents' income was 100% more than the total sample for 1985–1991, the data for 1991 provided in yearbooks 1993 and 1995 differed from each other.

Next, the standards vary by year. For example, before 1982, the statistics were based on registered residences: from 1982 to 1989, the numbers were derived from the third and fourth census; for 1990–2000, the data were adopted from the fifth census conducted in 2005; and after 2001, the data were taken from the sample surveys. While the fifth census considered all those who lived in a city for more than six months as city residents, the fourth census regarded those who lived in a city for more than a year as city dwellers. According to China's household registration system, city residents are those who have registered residences in cities regardless of where they live. By contrast, the statistical yearbooks also include the following as city residents: those with a registered city residence but living in the countryside for more than a year and those from the countryside and who do not own an urban residence but have been living in a city for a long period. Following is another example of varying statistics: in 1998, the percentage of the urban population was 33.35% as per the fifth census, 30.4% according to the fourth census, and 24.7% based on the registration system.

Further, the approach to a census differs between China and other countries in the world, which often goes unnoticed. For instance, households in China are required to record their total annual income and expenses, whereas many other countries use weekly, bi-weekly, or monthly income statements and multiply it by 52, 26, or 12 to obtain annual data (Gibson, Huang and Rozelle, 2001). Comparatively, a one-year model will decrease fluctuations in income; for example, one month's shortage will be compensated by another month's windfall. Therefore, it is reasonable to state the Gini ratio for China is underestimated, compared with those of many other nations (Deaton, 1995).

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

2.2 The Gini coefficient based on other data sources

Using adjusted data from urban and rural surveys for 1988 and 1995, Khan and Riskin (2001) and Zhao, Li and Riskin (1999) showed that income inequality continues to rise in China, a conclusion that has attracted wide attention. The sample for 1998 included 10,258 residences from 28 provinces, while that for 1995 comprised 7,998 residences from 19 provinces. By contrast, the corresponding data from the National Statistics Bureau were 51,352 and 34,739 residences. Although the sample data were considerably smaller than those of the NBS, the estimations were far more accurate. Nevertheless, it is difficult to estimate long-term fluctuations in income inequality using two years of data. The State Council Research Group (1997) also provided a few years' estimations of the nationwide Gini ratio. However, accurately reflecting the general trend of income inequality in China using limited data can be challenging. Thus, to accurately calculate the Gini coefficient using Equation (1), access to the original household survey data is warranted. Unfortunately, we are far from reaching this ideal status with the current data sources and there remain several technical difficulties.

2.3 Different methods

There are numerous shortcomings in the published data and obtaining the original data source used by NBS can be difficult. This has compelled researchers to explore other ways to resolve the problem. At present, researchers decompose the Gini coefficient to calculate the Chinese Gini ratio. The decomposition of the Gini index is not a new topic in econometrics and doing so using different income resources is a well-known practice. However, decomposing the Gini index for different groups remains challenging. Khan and Riskin (2001), for example, did not consider it appropriate to decompose the Gini coefficient on the basis of different groups. Bhattacharya and Mahalanobis (1967) were the first to conduct research on the topic. Pyatt (1976) divided the Gini index as per various income levels using game theory. Adopting matrix and covariance, Mookherjee and Shorrocks (1982), Shorrocks (1984), Lambert and Aronson (1993), and Cowell (2000) attempted to develop new methods to decompose the Gini ratio. However, their methods were deemed rather complex. Yao (1997) provided a comparatively simple method to decompose the Gini ratio. In brief, the Gini ratio for the entire population could be decomposed into the following (Yao, 1997):

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

where G_g is the Gini ratios for different groups, G_i is the Gini ratio within the i^{th} group, I_i is the proportion of group i^{th}'s income to total income, P_i is the proportion of the i^{th} group's population to total population, and the value of $G(f)$ depends on the extent of overlap among the different groups. If there is no overlap, $G(f)$ is 0 (Mookherjee and Shorrocks, 1982; Shorrocks and Wan, 2005). Milanovic (2002) indicated that the global Gini coefficient in 1993 was 0.578 and $G(f)$ was 0.068. This equation explains why the Gini ratio is underestimated based on the statistical yearbooks. In the *China Statistical Yearbook*, the income of city and rural residences were ranked from low to high, and therefore, $G(f)$ within either the city group or rural group was 0. However, when using data from the *China Statistical Yearbook*, there is no way to calculate the Gini ratio within each group. In other words, $\sum_{i}^{n} P_i I_i G_i$ of Equation (2) is ignored, and thus, the final results are eventually underestimated.

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

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

where δ and β are the results of the population proportion of the urban and rural areas (P_u and P_r) multiplied by the income proportion of the urban and rural areas (I_u and I_r), that is, $\delta = I_u P_u$ ($\beta = I_r P_r$). The decomposed nationwide Gini ratio includes the intra-rural Gini ratio, intra-urban Gini ratio, and Gini ratio between urban and rural areas, whose coefficients are δ, β, and 1, respectively. Equation (3) clearly demonstrates the contributions of each segment to the nationwide Gini ratio. If there is an overlap between urban and rural area income, for example, the income of certain urban residences is lower than the higher incomes of the rural residences, Equation (3) becomes

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

where G_0 stems from the overlap term of rural and urban income distribution. Theoretically, if a rural residence's income is higher than that of an urban residence, $G_0 > 0$, In this case, Equation (3) will underestimate the nationwide Gini ratio if there is an overlap of rural and urban income distribution. To consider an extreme example, when the population and income of a rural area is equal to that of the urban area, the Gini index from Equation (3) will be underestimated by 50%. Therefore, the value of G_0 is 0–0.5 G. In fact, Equation (4) represents a specific case of Equation (2). Chen, Hou and Jin (2008) calculated the Chinese Gini ratio but failed to consider the influence of G_0, and thus, presented an underestimated result.

Both local (Xiang, 1998; Chen and Zhou, 2002) and foreign (Sundrum, 1990; Yang, 1999) studies calculates the nationwide Gini ratio using

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

where Y_u, Y_r and Y are the annual average income of urban residents, rural residents, and the entire population, respectively. Chen and Zhou (2002) and Xiang (1998) used this formula without considering an income overlap between urban areas and rural areas, thus ignoring G_0 and producing an underestimated result.

$\frac{P_u P_r \left| Y_u - Y_r \right|}{Y}$ is the Gini ratio between rural and urban areas (derivations omitted). Since the values of Y_u and Y_r differ by society, we adopt its absolute value. Although equations (4) and (5) appear different, they are essentially the same following a transformation (derivation omitted) with one exception: Equation (4) offers a more visual representation of the contributions from the intra-rural Gini ratio (G_r), intra-urban Gini ratio (G_u), and Gini ratio between urban and rural areas (G_{ur}). As per Equation (4), the contributions from the intra-rural Gini ratio (G_r), intra-urban Gini ratio (G_u), and Gini ratio between urban and rural areas (G_{ur}) are $\beta G_r/G$, $\delta G_u/G$, and G_{ur}/G, respectively.

Cheng (2006, 2007) proposed new methods to resolve the problem of underestimation due to an overlap between a high rural income group and low urban income group. Chen's method requires the assumption of an income distribution function for urban and rural residences. However, there are an insufficient number of samples in the statistical yearbooks to do so. Such a limited number of groups rarely derives a reliable function for Chinese income distribution as a whole (Huang, 2007; Wang, 2007). According to Cheng, the contribution of the Gini ratio between urban and rural areas (G_{ur}/G%) in 1990 was 22.24%, which considerably differs from the 53% proposed in Chen and Zhou (2002). Lin, Cai and Li (1998) used Theil entropy index to examine income disparity among rural areas, across urban areas, and between urban and rural areas and found that the difference between urban and rural areas is almost always about 50% and has the highest influence on the overall effect. Wei (1996) showed that such contributions for 1985–1995 average about 51%. Khan and Riskin (2001) further stated that the Chinese Gini ratio is higher than that of rural and urban areas, indicating the importance of the income gap between the urban and rural areas. Wan (2004), Chen, Jin and Tang (2005), and Li, Sicular and Gustafson (2008) offer similar results.

Hong (2008) adopted the method of decomposing the Gini coefficient to address the overlap of high-income rural residents and low-income urban residents. This method decomposed the income inequality of the entire country on the basis of four groups: rural areas, urban areas, rural subgroups, and urban subgroups. However, it did not consider the income disparity between urban and rural areas. In sum, accurately calculating G_0 is a key element in estimating the Chinese Gini coefficient.

Few other studies calculated the Gini index adopting other perspectives. Most studies showed that Chinese income inequality from 1983 to 1984 was the lowest, particularly since 1978. After applying a weighted combination of data from Zhao and Li (1997), Huang and Xi (1999) concluded that the Chinese Gini ratio increased from 1978 to 1984; moreover, their nationwide Gini coefficient is considerably lower than that in other major studies. Some scholars presented values that were even lower than that of the intra-rural Gini ratio provided by the DRSES. Hu (2004) used a simple method with certain demands for data resources, although it provides only a generic estimation.

From the above review, it is fairly clear that the literature on the decomposition of the Gini coefficient lacks an effective way to calculate G_0 using Equation (4). In the next Section, we focus on exploring the influence of G_0 on nationwide Gini coefficients and the main characteristic of G_0.

3. Influence of G_0 on nationwide Gini coefficient

As mentioned above, Equations (4) or (5) can be employed to estimate the nationwide Gini ratio; however, how to calculate G_0 is a major problem. If we adopt Equation (3) to estimate nationwide Gini coefficients, we must omit G_0 and bear the risk caused by G_0. For example, if the population as well income distribution of a city and country is same, G_0 will account for 50% of the overall Gini ratio. Hence, it is necessary to discuss the features of G_0.

Early studies are often deemed unclear on the influence of G_0, which was generally termed residual or income overlap. Bhattacharya and Mahalanobis (1967) called it the concentrated area of income, whereas Mookherjee and Shorrocks (1982) labelled it a tricky interactive effecting equation that is impossible to accurately calculate. Lambert and Aronson (1993) believed that the residual was a result of the common effects among and between groups, which presented the overlap of income distribution in various groups. Cowell (2000) proposed similar views. The abovementioned studies suggested that the decomposition of the Gini ratio was influenced by G_0. According to Li (2005), the higher the earnings gathering in different groups, the lower is the overlap of individual income between them. Lambert and Decoster (2005) offered the following mathematical expression for G_0:

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

where $F_u(x)$ and $F_r(x)$ are the income distribution function of urban residents and rural residents and μ is the average income of residents across the country. If all Gini coefficients are decomposed according to rural and city residents and their income distribution follows a certain statistical distribution, the overlapping degree of income for urban and rural residents is closely related to the proportion of income for urban and rural residents and that of their population. Given the limitation of current data sources in China, the calculation of overlapping income between rural and city residents (G_0) is yet to be found in the literature.

3.1 Distribution function of income

To study the influence of G_0 on the calculation of nationwide Gini coefficients, we first fit the distribution of China’s urban and rural residents’ income. However, before doing so, it is important to identify the type of distribution function we can use to depict residents’ income distribution. Steyn (1950) claimed that the income distribution of individuals with the same characteristics, such as rural or urban residents, can be soundly described using lognormal distribution. Balintfy and Goodman (1973) emphasised that income distribution is generated through a special random process which can be explained using lognormal distribution. Cheng (2005) pointed out the universality of the lognormal distribution among the distribution of socio-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 conducted a study in 2006 and proved that income distribution was the lognormal distribution using data of nearly 40 years of residents' income in both developed and developing countries (Lopez and Servén, 2006). Souma (2000) examine Japanese residents' income for 1887–1998 and highlighted 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 for 1970–2003 based on the different development levels of economies and regions. Hong and Li (2006), on the other hand, presented a different view: income distribution can be generally classified into pyramid, dumbbell, and olive types, all of which may not follow lognormal distribution. In reality, however, the pyramid and olive types are rarely found in China. For example, if the income distribution is an olive type, its mode is close to its mean, which does not correspond to China's resident income distribution. There is a wide gap between urban and rural income because of China's dual social structure. In fact, the income distribution of all residents more closely resembles the dumbbell type.

The literature has few empirical studies on Chinese residents' income distribution function. As mentioned, the key problem is the lack of raw household survey data. The *China Statistics Yearbook* provides limited grouping data and thus, is not reliable to derive a distribution function. Therefore, at present, a key obstacle in examining Chinese income disparity is to determine an income distribution function using the current data of resident income.

This study attempted to resolve this problem by adopting suitable distribution functions. Our literature review revealed about 15 distribution functions which can be used to describe resident income distribution. Of these, we selected four commonly used distribution functions:

lognormal, gamma, exponential and Dagum distribution. Gibrat (1931) posed that the income distribution can be fitted well using lognormal distribution. Relying on per capita disposable income, Wang (2009) studied the income distribution of rural residents and concluded that lognormal distribution had the best fitting effect. Mao et al. (2009) adopted Gamma distribution to fit urban Chinese households' income from 2005 to 2007. Dagum (1977) proposed a multi-parameter distribution function, the Dagum distribution, which later became widely accepted.

It is almost impossible to precisely estimate the parameters of multi-parameter distribution functions using limited information (Chen, 2016). Therefore, we apply Dagum only as a multi-parameter distribution function and exclude the rest: for example, the Singh–Maddala distribution (SM), beta distribution of the second kind (B2), generalised beta distribution of the second kind (GB2), generalised beta distribution (GB), and exponential generalised beta distribution (EGB). Dagum's parameters can be adopted in the context of our study if relevant indexes, for example, the Gini ratio, mean, or median, are provided for the samples.

We examined residents' income distribution function using urban household survey data provided by the Sichuan Statistical Bureau in 2008, which includes 10,925 household samples. The distribution of random variables can be described by their probability density function (PDF) or cumulative distribution function (CDF). The PDF for the abovementioned four distributions can be indicated as follows:

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

Among the four CDFs, the CDF of exponential distribution is monotonically decreasing, that of the lognormal distribution is slightly to the right, and the distribution configuration of the gamma and Dagum distributions significantly change with the parameters.

First, we compared the above four distributions on the basis of urban residents' disposable income data for 2008, which were provided by the Sichuan Statistics Bureau. Given the inconsistencies between the actual population proportions and the samples' proportions for different areas, we determined the weight of every sample according to real population in all urban areas. Owing to the existence of weight, direct K-S test is not suitable. If we revert the data according to the sample weight, because the weight should be accurate to the fourth decimal place, this would increase the sample by 10,000 times. Given the resultant size of the sample, the software would be unable to process it and thus, we adopted another method.

The main objective of this study is to identify the function which has the best goodness-of-fit. To do so, we compared the distribution function with one based on theory using empirical data. Taking exponential distribution as an example, for any fractile x and a certain parameter λ, we can write the theoretical distribution function of such a fractile as follows:

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

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

$$y(n) \leq x, y(n+1) > x . \tag{9}$$

Therefore, the distribution function of the empirical data is

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

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

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

By adjusting λ, we can find a λ to minimise the error, and then, treat this λ as our estimate parameter. Similarly, we can estimate the parameters and errors of the three other distributions. The results are as follows.

[Insert Table 1 about here]

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

To confirm this result, we adapt a natural logarithm to income, and then, create a histogram and normal Q-Q plot (see Figure 1). Figure 1 (left) is the histogram which adapts the logarithm of the original data and it is largely similar to the histogram of normal distribution. Furthermore, we compared its real quantile with the theoretical quantile of normal distribution. In the identification process, the figure should be a line passing through the origin and the slope should be at 1. In Figure 2 (right), the data were almost identical to the theoretical lognormal distribution. In addition, we accessed urban and rural household samples from a 2008 survey conducted in Chengdu. When we mixed all household survey samples, the samples did not obey the lognormal distribution but resembled a dumbbell type.

On the other hand, if we study urban and rural resident income distribution separately, we find that they all obey lognormal distribution.

[Insert Figure 1 about here]

As previously mentioned, according to the statistics yearbook, only grouped household survey data can be found in current published yearbooks. The grouped data are, however, over-concentrated, making it difficult to precisely predict the Gini ratio. To solve this problem, we must generate household income data using current data sources, which will then replace raw household survey data. Doing so will not be difficult if the variance in and expectation from raw household survey data are known. Here, the expectation-maximisation (EM) algorithm is employed to estimate two parameters: variance and expectation.

The EM algorithm as an iterative method (Deng and Yang, 2004) is mainly utilised to identify the mode of posterior distribution, that is, maximum likelihood estimates. An iteration comprises two steps: the E-step (expectation) and M-step (maximisation). The fundamental objective of the EM algorithm can be manifested as follows. With θ as the unknown parameter, we can use Matlab to iterate the two steps until $\|\theta^{i+1}-\theta\|$ is minimised. The advantages of the EM algorithm are simplicity and stability. Suppose the i^{th} resident's income x_i (i = 1,2,..., n) is an independent and identically distributed random variable and obeys the lognormal distribution with μ and δ as parameters. Then, the density function is

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

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

Controlling for the amount of random numbers on the basis of the real urban and rural population, and merging the urban and rural income distributions to derive the national resident income distribution, can help calculate the national Gini coefficient, intra-rural Gini coefficient, intra-urban Gini coefficient, and urban–rural Gini coefficient. G_0 can be obtained either by using Equation (6) or through the difference between equations (3) and (4). Because $F_u(x)$ and $F_r(x)$ are so complex, it is difficult to calculate G_0 using Equation (6). Using this method, we analysed statistical data for 2005, which showed that the Gini ratio of urban and rural resident income is relatively close to the intra-urban or intra-rural Gini ratios estimated using data from the USESC and DRSES. Here, our focus is to explore the change trend of G_0 on the basis of these findings instead of accurately predicting the national Gini index. Moreover, it is noteworthy that the 20 grouped data for urban resident income can only be traced back to 2005 and therefore, we cannot fit the distribution of the urban resident income before 2005.

In sum, from the previous results, real data for 2005, and controls for certain variables, we observed how changes in the proportion of urban population and the urban–rural per capita income ratio influence G_0. The following results are acquired through Matlab programming.

3.2 Main features of G_0

Figure 2 (left) illustrates the 'first rising then descending' trend of G_0's share in the national Gini coefficient if the share of urban population increases from 5% to 90%. As the urban population rises from 5% to 90%, G_0 increases from 0.0071 to 0.0214 and then, decreases to 0.0049. The corresponding proportions in the national Gini coefficient are, respectively, 1.68%, 4.54%, and 1.35%, with 30% of the urban population contributing to the highest (4.54%) value.

[Insert Figure 2 about here]

Next, we examined how changes in urban–rural resident income affect G_0. Figure 2 (right) demonstrates that as the average income ratio changes from 0.1:1 to 4.3:1, the change trend for G_0 first rises and then descends, which takes a different form in this instance. When the urban and rural average income ratio is 1:1, G_0 is maximised—it accounts for nearly half of the Gini coefficient of all residents' income. However, with a rising urban–rural resident income, the proportion of G_0 in the national Gini coefficient increasingly descends. Because the ratio of urban–rural average income is greater than 2.5:1, the proportion of G_0 is less than 10%. Here, we explore changes in G_0 as a result of simultaneous changes in urban population and the urban–rural average income. Here, suppose the proportion of urban population changes from 10% to 90% and the urban–rural income ratio changes from 1.8:1 to 3.9:1.

[Insert Figure 3 about here]

As the 3D graphs (Figure 3) depict, when the proportion of urban population is about 40% and the urban–rural income ratio is 1.8:1, G_0 is maximised. On the other hand, when the former is about 90% and the latter is 3.9:1, G_0 is minimised. Therefore, the overall trend of the G_0 change is as follows: when the urban–rural per capita income ratio (R) increases, G_0 descends and when the proportion of urban population increase, G_0 first rises and then descends. In addition, we can analyse the contribution rate of G_0 to residents' income. Figure 3 (right) demonstrates that when the urban population is about 40% and the urban–rural per capita income ratio is 1.8:1, G_0 is maximised, while the national Gini coefficient is considerably low. As a result, the contribution rate of G_0 to the nationwide Gini ratio is up to 18.19%. By contrast, as the urban population is 90% and 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%. According to the *China Statistical Yearbook 2006*, the proportion of China's urban population in 2005 is 43%; the urban–rural average income ratio is 3.2237:1; and the corresponding G_0 is 0.0189, accounting for 4.1% of the national Gini coefficient. Assuming all other variables are constant, an increase in the proportion of the urban population will cause G_0/G to descend.

Note that, however, the proportion of urban population and the urban distribution variance of the urban and rural resident income are fixed. Thus, to calculate the national Gini coefficient since China's economic reform, it is necessary to measure the variance and expected value of the distribution of urban and rural residents' annual income. This appears to be the only method to accurately calculate G_0.

In the next section, we introduce methods to measure the variance and expected value of urban and rural residents' income distribution, calculate the corresponding G_0, and finally, estimate the national Gini coefficient.

4. Estimation of national Gini coefficient

4.1 Method

First, using Equation (3), we calculate the national Gini coefficient but exclude G_0. We will use the intra-urban and intra-rural Gini coefficients provided by the USESC and the DRSES. Since their surveys can ensure initial data, the resultant Gini coefficient estimated can be deemed reliable. However, the USESC has not provided the intra-urban Gini coefficients since 2011; therefore, we can only estimate the Chinese Gini ratio from 1978 to 2010. The Gini coefficient between urban and rural areas (G_{ur}) is easy to calculate and not affected by intra-group income inequality generated by grouped samples. Thus, we can estimate the national Gini coefficient by excluding G_0. Second, we will introduce a new statistic method to calculate G_0. Finally, we will use Equation (4) to calculate the national Gini coefficient.

First we illustrate the application of the statistic method to calculate G_0. Suppose μ and δ are the mean and variance of lognormal distribution. Adopting the numerical calculating method, Cheng (2005) obtained the Gini coefficient corresponding to a different δ. Hong and Li (2006) proved that, for any income variant x, if Lnx ~ N (μ, δ^2), then the Gini coefficient is

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

In Equation (13), for any δ, the Gini coefficient can be easily calculated using the table of the standard normal distribution function. By contrast, here μ and δ are calculated using the intra-urban Gini ratio, intra-rural Gini ratio, as well urban and rural per capita income provided by NBS. Since $Y = \exp(\mu + \frac{\delta^2}{2})$, where Y is the known urban and rural average income, Equation (14) can be obtained as follows:

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

The intra-urban and intra-rural Gini ratios for 1978–2010, as provided by the USESC and the DRSES, range from 0.15 to 0.385 (see Table 2), with δ ranging between 0.25 and 1. Through Matlab programming, 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. In addition, our programming can ensure a corresponding δ after inputting the Gini ratios in urban and rural areas. Using δ and the known urban and rural per capita income (Y), μ can be calculated using Equation (14). In this way, we can obtain the income distribution functions of urban and rural residents for 1978–2010, separately. Then, with Matlab, we can generate the random numbers that obey lognormal distribution using parameters μ and δ, which represents individual residents' income and approximately fit real income data. By controlling for random number rates on the basis of real urban and rural population, sample data for national resident income can be obtained. Then, with Equation (1), the national Gini coefficient (G) and national Gini coefficient ($G^{'}$) excluding G_0 can be easily calculated. Then, since $G_0 = G - G^{'}$, G_0 can be estimated.

All of the above calculations are realised by inputting the intra-urban and intra-rural Gini coefficients of a specific year, urban and rural per capita income, proportion of urban population, and national Gini coefficient and G_0 can be directly outputted. In the actual computations, there are a total of 10,000 Matlab-generated random numbers that obey lognormal distribution. These numbers are distributed on the basis of the real urban–rural population ratio for a specific year. G_0 is the mean number in the 100 times of simulation. To check the convergence of the calculation results, we first derive $G^{'}$ for 1978–2010 using random numbers. Then, with Equation (3), we obtain $G^{'}$ for 1978–2010 using the intra-urban and intra-rural Gini coefficients, urban and rural per capita income, and urban–rural population ratio provided by NBS and *China Statistical Yearbook 2011*. We find an identical $G^{'}$ despite using two methods, which means following 100 simulations, G_0 steadily converges to a specific datum.

4.2 Advantages and disadvantages of method

First, the method can overcome the present restriction of data sources and ensure the accurate estimation of China's Gini coefficient. Using the initial data from the survey on the residents, the USESC and the DRSES allow for a more accurate estimation of the Gini ratios in urban and rural areas than the statistical yearbook. Meanwhile, the G_{ur} calculation is not restricted to the limited grouped data from the statistical yearbook. Although the proportion of G_0 in the national Gini ratio is not high, the current grouped data in the statistical yearbook cannot accurately reflect the overlapping degree of urban–rural residents' income. However, this problem can be solved using the statistical method proposed in this study. The method can overcome the current deficiency in data resource needed to calculate the Gini coefficient of Chinese residents' income.

Second, the method can be adopted to quantitatively analyse the structure of the national Gini coefficient. The decomposed Gini coefficients include the intra-urban and intra-rural Gini coefficients and the Gini coefficient between the urban and rural areas. Thus, we can quantitatively estimate the influence of each element on national income inequality and identify the leading factor.

Finally, it can be used to perform continuous calculations of the national Gini ratios since the Chinese economic reform period. Given that NBS and *China Statistical Yearbook 2011* offer data on intra-urban and intra-rural Gini coefficients, urban and rural resident per capita income, and urban–rural population ratio, using our method can prove convenient when calculating using data since 1978. As a result, it makes it possible to analyse the change trend for income inequality since the economic reform.

Despite the abovementioned advantages, the disadvantages of this approach are obvious. First, if the USESC and the DRSES do not provide intra-rural and intra-urban Gini coefficients, our method will be unable to derive rural and urban income distribution functions. While there are many distribution functions, we cannot deduce the parameters of a majority multi-parameter distribution functions using limited information for income distribution (e.g. Gini ratio, mean, or median). Nevertheless, in most instances, the goodness-of-fit of multi-parameter distribution functions are better than two-parameter distribution functions.

In summary, contributions of the new approaches are as follows: 1. We overcome the limitations of data sources that block the accurate estimation of the Gini ratios; 2. we calculate the Chinese Gini ratios from 1978 to 2010, which are not provided in relevant studies; and 3. the methods allow us to investigate the trend of Chinese Gini ratio.

4.3 Results

According to Table 2, the national Gini coefficient increases from 0.3043 in 1978 to 0.4519 in 2010 and the national Gini ratio of 2010 is 1.49 times greater than the value reported for 1978. G_r, G_u, and G_{ur}, respectively, increase to 78%, 106%, and 46% for 1978–2010, during which the contribution rate of G_0 to the Gini ratio of national residents' income is between 0.46% and 7.55%. In general, the contribution rate of G_0 is small. However, in 1983 and 1985, the rate of contribution for G_0 exceeded that for G_u to the national Gini ratio. More specifically, the urban–rural income gap becomes the smallest in the mid-1980s, that is, since China's economic reform. As a result, the urban and rural income considerably overlaps and this renders G_0 with a rather large value. As discussed in Section 3, with the increase in the proportion of urban population, the contribution rate of G_0 to the Gini coefficient of the national residents' income depicts a 'first rising and then descending' trend. This trend is in accordance with the change trend for G_0 from the early 1990s to 2010. In Table 2, G_r (intra-rural Gini coefficient) in Column (1) and G_u (intra-urban Gini coefficient) in Column (2) are directly adopted from *Chinese Income Distribution Yearly Report* (2011) which was edited by the Employment and Income Distribution Department of Development and Reform Commission of China, in line with this book, G_u and G_r from 1978 to 2010 were provided by the USESC and the DRSES.

[Insert Table 2 about here]

Between 1978 and 2010, the contribution rate of the intra-rural Gini coefficient to the national Gini coefficient decreases from 36.7% to 9.92% and that of the intra-urban Gini coefficient to the national Gini coefficient increases from 3.39% to 27.84%. Following this trend and 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 rapidly increased, 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 shown in Table 2, G_{ur} remains the determining factor of the national Gini coefficient, and in 2010, it accounted for 58.33% of the national Gini coefficients. In addition, the two present the same change trend. Therefore, considering the dominating effect of G_{ur} on the change in the national Gini coefficient and the same change trend, we now shift our attention to the urban–rural income gap.

Our estimation results might be lower than those offered by other researchers, such as the *Chinese Household Financial Survey* (CHFS). As previously mentioned, such gaps can be mainly attributed to different sampling methods. For instance, the Gini ratio obtained for households on the basis of total annual income is probably smaller than that derived from weekly, bi-weekly, and monthly data, because the dispersion degree of the latter is likely to be greater than the former. However, current studies have neglected this issue.

[Insert Table 3 about here]

As mentioned above, NBS has presented the nationwide Gini ratios since 2013, which are based on the data of USESC and DRSES. Comparing our estimation with NBS' results (see Table 3), we found that our calculation included approximately 2.9%-6.4% underestimation. NBS admitted that the real income of urban high-income group was underrated, and therefore NBS adjusted the income of the urban upper crust to correspond with the information of individual income tax. This is the main reason for the differences between the results of NBS results and our estimation. However, NBS' data can only be traced back to 2003; thus, it does not provide the nationwide Gini of the period when reforms and open-door policies were adopted. Meanwhile, the results of nationwide Gini coefficients are our only barriers in investigating the structure of nationwide Gini and its changing tendency.

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

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

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

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

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

With an increase in urban population (P_u), the Gini ratio between urban and rural areas becomes larger. When G_{ur} reaches its max (G_{urmax}), a further increase in P_u will cause G_{ur} to descend. As a result, we observe a U-shaped curve. 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, that is, with an increase in R, the influence of R on G_{ur} decreases. Therefore, urbanisation is crucial to the decline in urban–rural income inequality.

The above discussion demonstrates that the key factor determining the trend of the Chinese Gini coefficient is the Gini coefficient between urban and rural areas. Therefore, it is necessary to explore the effect of the urbanisation process and changes in urban–rural per capita income on national income inequality.

5.1 Influence of urbanisation process

Adopting data on the intra-urban and intra-rural Gini coefficient by the USESC and the DRSES in 2005 and on the urban and rural population proportions as well urban and rural per capita income by the *China Statistical Yearbook* 2006, we can calculate the main parameters of the income distribution of urban and rural residents for 2005 using Equation (14). The income distribution functions of urban and rural residents can be used to explore the influence of the urbanisation process and changes in the per capita income ratio on national income inequality. First, without changing the values of μ and δ, this influence can be observed by increasing the proportion of the urban population from 5% to 90%.

[Insert Figure 4 about here]

Figure 4 demonstrates the mean number in 100 repeated simulations of the 10,000 computer-generated random numbers using the known distribution function. Figure 4 (left) shows that as the proportion of the urban population rises from 5% to 90%, the national Gini coefficient increases from 0.4230 to 0.4755 and then declines 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 in the case of urbanisation: income inequality first rises and then descends. The proportion of urban population corresponding to the highest national Gini coefficient is about 26% and that of real urban population in 2005 reaches 43%, which means that urbanisation will continue to contribute to national income equality.

In line with Lewis' (1954) dual economy theory, urban industrial sectors may gain the highest profits from society during the initial period of industrialisation. As a result, income inequality between urban industrial sectors and rural traditional sectors is likely to increase. Further, the growing rate of rural population transferring to urban areas is likely to deplete surplus rural labour. Consequently, real rural income and a society's average income will be enhanced, which will certainly diminish the urban–rural income inequality. Drawing on Lewis's dual-sector model to review the economic development of developed countries, Kuznets (1955) held that the evolution of income inequality is an invert U shape, that is, it first rises and then descends. Therefore, urbanisation exerts a critical effect on national income inequality.

5.2 Influence of urban-to-rural per capita income ratio

According to *China Statistical Yearbook* data for 1978–2010, in 1983, the urban–rural per capita income ratio was at its lowest, 1.8225:1 and in 2009, it reached its highest, 3.333:1. Using 2005 data and without changing parameter δ, we change the ratio from 1.5:1 to 4:1 to observe changes in national income inequality. According to Figure 4, when the ratio changes from 1.5:1 to 4:1, the national Gini coefficient increases 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 in the urban–rural per capita income ratio on the urban–rural Gini coefficient exceeds that on the national Gini coefficient. The urban–rural Gini coefficient's share in the national coefficient also increases from 26.6% to 63.6%. We also discover that a growing urban–rural per capita income ratio may deteriorate income inequality; however, this influence has a convergent tendency. As previously discussed, 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. In the case of expanding indicators of the urban–rural per capita income, the following can be clearly observed: G_{ur} is a curve with a crest in the upper left field, that is, with the continuous increase in the urban–rural per capita income ratio, the growth rate of the urban–rural Gini coefficient has a downward tendency.

5.3 Co-influence of population and urban-to-rural per capita income ratio

In this subsection, we explore how the national Gini ratio changes if the proportion of urban population and urban–rural per capita income ratio simultaneously change. Here, we change the urban population share from 10% to 90% and the urban–rural per capita income ratio from 1.8:1 to 3.9:1. Figure 5 (left) demonstrates that, with an increase in the proportion of urban population, the national Gini ratio first rises and then descends, and with the rise of the urban–rural per capita income ratio, the proportion of urban population corresponding to the highest value of the Gini ratio for national resident income also increases. We take the year 2005 as an example. When the proportion of urban population reaches 26%, a further increase in urban population will diminish 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 accounts for 30%. According to the *China Statistical Yearbook 2011*, the proportion of urban population in 2010 is 49.95% and the urban–rural per capita income ratio is 3.2285:1. Clearly, further urbanisation will contribute to an improvement in the national income distribution.

[Insert Figure 5 about here]

Given the key role of urban–rural income gap in a national income gap, here, we explore changes in the share of urban–rural income inequality in national income inequality. Figure 4 demonstrates that, the share of G_{ur} in the national Gini coefficient increases with a rise in the urban–rural income ratio. Meanwhile, if the proportion of urban population rises, the share of G_{ur} in the national Gini coefficient first rises and then descends. In this case, when the urban population accounts for 35% and the urban-to-rural per capita income ratio is 3.9:1, G_{ur} reports the highest share of 65.39% of the national Gini coefficient. By contrast, when the urban population has a share of 90% and urban–rural per capita income ratio is 1.8, G_{ur} records the smallest share of the national Gini coefficient, 12.18%. Even if the urban–rural per capita income ratio remains at 3.9:1 and the proportion of the urban population is as high as 90%, G_{ur} accounts for only 19.75% of the national Gini coefficient. Therefore, urbanisation is beneficial to decreasing the share of G_{ur} in the national Gini coefficient.

6. Conclusions and policy suggestions

Present data sources are responsible for the on-going debates on China's income Gini index. To overcome this limitation, much of the current research is attempting to make a breakthrough by improving calculation methods, especially by decomposing the national Gini ratio by urban and rural areas. However, there is an overlap between urban and rural resident income, which can affect the accuracy of the calculation of the national Gini coefficient. Thus, this study examined the characteristics and influences of G_0 by applying the two-step EM algorithm to fit China's urban and rural resident income distribution for 2005. Accordingly, we compiled a program to analyse how changes in population and urban–rural per capita income ratio affect the value of G_0.

Moreover, given the present condition of data sources, we calculated the Chinese Gini coefficient since China's economic reform using the statistical method. By decomposing the national Gini coefficient, we find that the Gini coefficient between urban and rural residents is the key factor affecting China's income inequality and urbanisation and changes in urban–

rural per capita income ratio mainly affect urban–rural income inequality. Using these results, we further explored the influence of change in urban population and urban–rural per capita income ratio on national income inequality. We found that urbanisation cause the national Gini coefficient and Gini coefficient between urban and rural areas to first rise and then descend. Although an increase in the urban–rural per capita income ratio enhance the degree of income inequality, its influence depicts a convergent trend. Therefore, diminishing urban–rural income inequality is mainly responsible for decreasing China's income equality and accelerating urbanisation can help reduce urban–rural income inequality.

There are two ways to reduce income equality between urban and rural areas. First is to increase farmers' income and thus, reduce the urban–rural income gap. For instance, China has implemented a series of reforms including the abolishment of agricultural tax; subsidies for direct grains, seeds, and agricultural machinery; and rural land circulation. Second is to accelerate urbanisation. According to NBS, the contribution rate of primary industries to the national GDP was a mere 4.6% in 2015, whereas the share of labour force in primary industries was 28.3% of national labour resources. Rural income will not significantly increase if such large numbers of farmers continue to rely on limited land resources over the next 20–30 years. Therefore, breaking the urban–rural dual social structure is difficult when only considering the question of how to increase rural income. At its third plenary session, the 17th Communist Party of China Central Committee stated the significance and urgency of breaking the urban–rural dual social structure through its judgements: China has arrived at an important period of breaking the urban–rural dual social structure and bringing in a new era of balanced economic and social development for urban and rural areas. In other words, the key to breaking the dual structure is to promote urbanisation. According to the national Gini ratio by NBS, the Chinese Gini ratio depicted a downward trend from 2008 to 2015, when urban population ratios reported a continuous increase.

Given the present conditions, 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 large-scale transfers of rural labour and other infrastructure pressure can be relieved by building city clusters and scattered small- and medium-sized towns. Further, opportunities to industrially upgrade the eastern district can help acquire investments from the eastern industries. In terms of industrial development, labour-intensive industries, especially the service industry, should be supported given the difficulties and obstacles in directly transferring education- and skill-limited rural labour to the modern sectors in cities. By contrast, with a low entrance threshold, the service industry can absorb a large number of rural labourers.

Another effective way to improve urban–rural income inequality is to transfer rural labour from remote less-developed areas in the middle and western parts of China to medium- and 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 a decline in rural-based poor population, the rural average income will increase and the urban–rural per capita income gap will diminish. Evidently, the smooth transfer of rural surplus labour relies on the capacity to create more employment opportunities derived from economic development. Economic growth cannot automatically solve China's income inequality, but it will be difficult to decrease income inequality without it.

References

Balintfy, J. L., and Goodman, S. S. 1973. "Socio-economic Factors in Income Inequality: A Log-Normal Hypothesis." *Zeitschrift fiir Nationalokonomie*, 33 (3): 389–402.

Bhattacharya, N., and Mahalanobis, B. 1967. "Regional Disparities in Household Consumption in India." *Journal of the American Statistical Association*, 62 (317): 143–161.

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

Chen, J. 2016. *On Income Distribution Functions and Decomposition of the Gini Coefficient as well as Their Application in China*. Beijing: Economics Science Press.

Chen, J., Jin, S., and Tang, X. 2005. "A New Method of Analyzing Urban-Rural Income Gap and an Analysis on 1986-2003 China's Urban-Rural Inequality." In *Studies on Income Inequality and Poverty in Transitional China*, edited Fang Cai and Guanghua Wan, 165-181. 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, July 2–4.

Chen, J., Dai, D., Pu, M., Hou, W., and Feng, Q. 2010. "The Trend of the Gini Coefficient of China." Brooks World Poverty Institute (BWPI) Working Paper 109. ISBN: 978-1-907247-08-8.

Chen, Z. 1999. *Reform, Development and Income Distribution*. Shanghai: Fudan University Press.

Chen, Z., and Zhou, Y. 2002. *Rethinking on Income Distribution in Reform and Development: China Has a Bipolar Economy?* Beijing: Economic Science Press.

Cheng, B. 2005. "Lorenz Curve and Gini Coefficient under Condition of Lognormal Distribution." *Journal of Quantitative and Technical Economics*, 22 (2): 127–135.

Cheng, Y. 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, Y. 2006. "Calculation and Decomposition of the Overall Gini Coefficient in Dual Economies." *Economic Research Journal*, 1 (1): 109–120.

Cowell, F. 2000. "Measuring Inequality." In *Handbook: Income Distribution*, edited by Anthony Atkinson and François Bourguignon. Amsterdam, New York and Oxford: Elsevier North-Holland.

Deaton, A. 1995. "Data and Econometric Tools for Development Analysis." In *Handbook of Development Economics*, edited by Jere R. Behrman and T.N. Srinivasan, 3. Amsterdam: Elsevier Press.

Deng, M., and Yang, Y. 2004. "Estimating Parameter in Lognormal Distribution from Grouped Data." *Systems Engineering –Theory Methodology Application*, 6: 553–556.

Employment and Income Distribution Department of Development and Reform Commission of China. 2011. *Chinese Income Distribution Yearly Report* (2011). Beijing: Economic Science Press.

Department of Rural 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 (4): 430–443.

Gibrat, R. 1931. *Les Inegalite´s E´conomiques*. Paris: Librairie du Recueil Sirey.

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 (3): 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, X. 2008. "A New Subgroup Decomposition of the Gini Coefficient." *China Economic Quarterly*, 1 (8): 307–342.

Hong, X., and Li, J. 2006. "A Further Study about some Problems of the Gini Coefficient." *Journal of Quantitative and Technical Economics*, 2: 86–96.

Hu, Z. 2004. "A Study of the Best Theoretical Value of Gini Coefficient and its Concise Calculation Formula." *Economic Research Journal*, 9, 60–69.

Huang, D., and Xi, Y. 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), 1–28.

Lambert, P., and Aronson, R. 1993. "Inequality Decomposition Analysis and the Gini Coefficient Revisited." *The Economic Journal*, 103 (420): 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, H. 2005. "A Research on Gini Coefficient Decomposition Analysis." *Quantitative & Technical Economics*, 3: 127–135.

Li, S. 2003. "Review and Prospect: Studies on China's Resident Income Distribution." *China Economic Quarterly*, 2 (2): 379–404.

Li, S. 2002. "A Further Discussion on Estimation and Decomposition of Gini Coefficient." *Economic Research Journal*, 5: 84–87.

Li, S., and Luo, C. 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, S., Sicular, T., and Gustafson, B. 2008. *Research on Income Distribution in China*. Beijing: Beijing Normal University Press.

Lin, Y., Cai, F., and Li, Z. 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.

Mao, L., Hao, C., Ping, X., Liang, S., and Zhang, H. 2009. "International Comparison of the Individual Income Tax Threshold and Effect Estimation on Raise of Minimum Threshold." *International Economic Review*, 6: 34–38.

Milanovic, B. 2002. "True world income distribution, 1988 and 1993: First calculation based on household surveys alone." *Economic Journal*, 112 (476): 51–92.

Mookherjee, D., and Shorrocks, A. 1982. "A Decomposition Analysis of the Trend in UK Income Inequality." *The Economic Journal*, 92 (368): 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." Accessed January 7 2010. http://www.stats.gov.cn/tjgb/index.htm.

Pyatt, G. 1976. "On the Interpretation and Disaggregation of Gini Coefficients." *Economic Journal*, 86 (342): 243–255.

Ravallion, M., and Chen, S. 2007. "China's (Uneven) Progress against Poverty." *Journal of Development Economics*, 82 (1): 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 (1): 59–82.

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

Souma, W. 2000. "Universal Structure of the Personal Income Distribution." Accessed January 8 2010. http://arxiv.org/PS_cache/cond-mat/pdf/0011/0011373v1.pdf.

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 State Council Research Group. 1997. "Analysis and Suggestions on Urban Residents Income Inequality." *Economic Research Journal*, 8: 3–10.

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, G. 2004. "Measure and Decomposition of Income Distribution: A Review on Research Methodology." *World Economic Papers*, 1: 64–69.

Wang, C., and Huang, S. 2007. "Gini Coefficient and Sample Information Content." *Quantitative and Technical Economics*, 2: 136–144.

Wang, Z. 2009. “The Determination of Poverty Level and Poverty in Rural Areas—Relying on the Fitting Method of Income Distribution Function.” *Northwest Population Journal*, 1: 1–6.

Wei, H. 1996. “The Income Gap of Residents in Different Regions in China and its Decomposition.” *Economic Research Journal*, 11: 66–73.

Xiang, S. 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 (1): 27–31.

Zhao, R., and Li, S. 1997. “The Reasons of Enlarged Income Disparity in China.” *Economic Research Journal*, 9: 19–28.

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

Acknowledgements

We appreciate the support of the National Natural Science Foundation of China (Nos. 71473203), the Program for the Key Program of National Social Science Fund of China (No. 15AJY005) and Sichuan Province Social Science High Level Research Team.

Table captions

Table 1. Parameters and goodness-of-fit error

Table 2. Gini coefficient and its decomposition, 1978–2010

Table 3. The different estimation of nationwide Gini ratio

Table 1. Parameters and goodness-of-fit error

Function	**Exponential**	**Lognormal**	**Gamma**	**Dagum**
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	0.039	0.0039	0.009	0.0093

Table 2. Gini coefficient and its decomposition, 1978–2010

Year	G_r	G_u	G_{ur}	G'	G_0	G	G_0/G (%)	$\beta G_r/G$ (%)	$\delta G_u/G$ (%)	G_{ur}/G (%)
1978	0.2124	0.160	0.1803	0.3029	0.0014	0.3043	0.46	36.70	3.39	59.25
1979	0.2245	0.160	0.1802	0.3073	0.0021	0.3094	0.68	37.42	3.53	58.24
1980	0.2407	0.160	0.1813	0.3144	0.0030	0.3174	0.95	38.19	3.67	57.12
1981	0.2406	0.150	0.1597	0.2940	0.0044	0.2984	1.47	41.12	3.66	53.52
1982	0.2317	0.150	0.1355	0.2665	0.0070	0.2735	2.56	43.64	4.02	49.54
1983	0.2461	0.150	0.1183	0.2570	0.0123	0.2693	4.57	47.67	4.03	43.93
1984	0.2439	0.160	0.1241	0.2587	0.0124	0.2711	4.57	44.73	4.81	45.78
1985	0.3072	0.190	0.1291	0.2927	0.0239	0.3166	7.55	46.92	5.21	40.78
1986	0.3042	0.190	0.1629	0.3182	0.0158	0.3340	4.73	40.69	5.69	48.77
1987	0.2889	0.200	0.1700	0.3159	0.0142	0.3301	4.30	37.70	6.49	51.50
1988	0.3053	0.230	0.1717	0.3263	0.0196	0.3459	5.67	37.34	7.38	49.64
1989	0.3185	0.230	0.1864	0.3410	0.0184	0.3594	5.12	36.06	7.52	51.86
1990	0.3099	0.230	0.1772	0.3319	0.0189	0.3508	5.39	36.32	7.64	50.51
1991	0.3072	0.240	0.1984	0.3491	0.0155	0.3646	4.25	32.66	8.32	54.42
1992	0.3134	0.250	0.2205	0.3690	0.0142	0.3832	3.71	29.98	8.86	57.54
1993	0.3292	0.270	0.2412	0.3928	0.0142	0.4070	3.49	27.92	9.67	59.26
1994	0.3210	0.300	0.2483	0.4004	0.0150	0.4154	3.61	25.79	10.98	59.77
1995	0.3415	0.280	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	0.2140	0.3699	0.0225	0.3924	5.73	26.43	12.74	54.54
1998	0.3369	0.300	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	0.2600	0.4173	0.0217	0.4390	4.94	18.59	17.64	59.23
2002	0.3646	0.320	0.2754	0.4321	0.0185	0.4506	4.11	16.44	18.50	61.12
2003	0.3680	0.340	0.2824	0.4467	0.0190	0.4657	4.08	14.68	20.35	60.64
2004	0.3692	0.325	0.2794	0.4394	0.0173	0.4567	3.79	14.24	20.71	61.18
2005	0.3751	0.320	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
2007	0.3742	0.340	0.2796	0.4478	0.0173	0.4651	3.73	11.39	24.77	60.12
2008	0.3776	0.338	0.2762	0.4455	0.0175	0.4630	3.77	10.98	25.59	59.65
2009	0.3850	0.335	0.2738	0.4447	0.0176	0.4623	3.80	10.45	26.52	59.23
2010	0.3783	0.330	0.2636	0.4342	0.0177	0.4519	3.91	9.92	27.84	58.33

Notes:
1. G_{ur} in Column (3) is based on the authors' calculation using data from the *China Statistical Yearbook of 20011* by NBS.
2. G' (national Gini coefficient) in Column (4) is based on the authors' calculation using Equation (1).
3. G_0 in Column (5) is derived from the authors' calculation using the statistical method.
4. G' (national Gini coefficient) in Column (4) is from the authors' calculation using $G = G' + G_0$.
5. Columns (7)–(10) are, respectively, the percentage shares of G_0, intra-rural Gini coefficient, intra-urban Gini coefficient, and Gini coefficient between urban and rural areas in the national Gini coefficient.

Table 3. The different estimation of nationwide Gini ratio

year	2003	2004	2005	2006	2007	2008	2009	2010
our estimation	0.4657	0.4567	0.4573	0.4624	0.4651	0.463	0.4623	0.4519
NBS' estimation	0.479	0.478	0.485	0.487	0.484	0.491	0.49	0.481

Figure captions

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

Figure 2. Main features of G_0

Figure 3. Change trend for G_0

Figure 4. Urbanisation process and urban–rural income gap

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

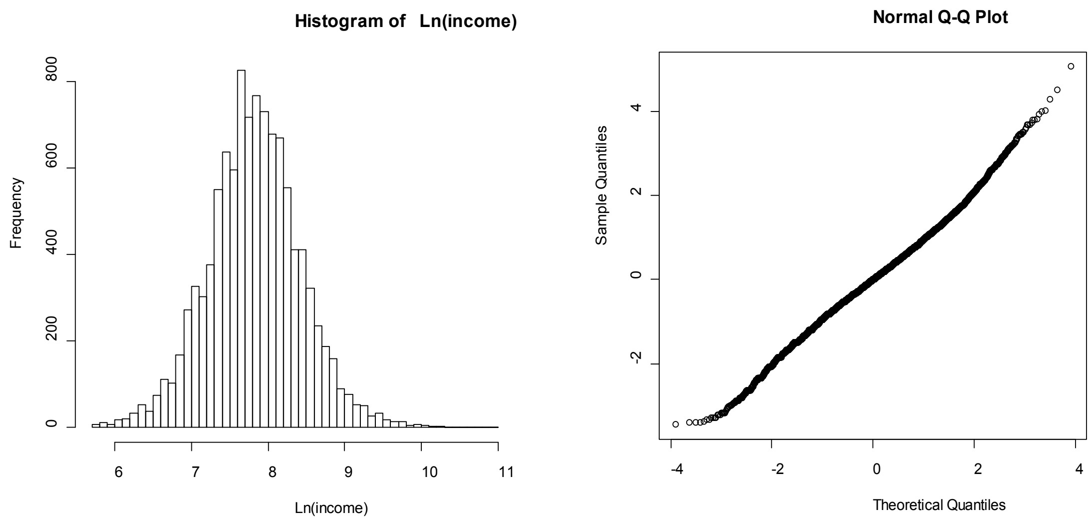


Figure 1. Histogram and normal Q-Q plot of income distribution

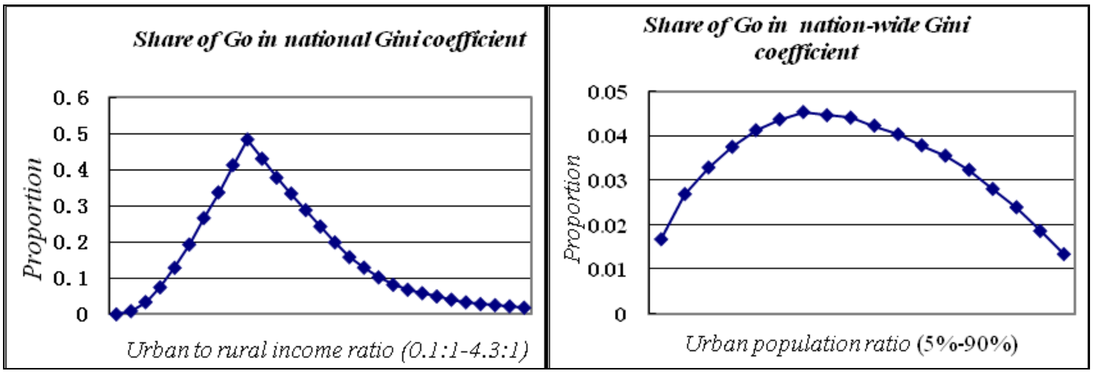


Figure 2. Main features of G_0

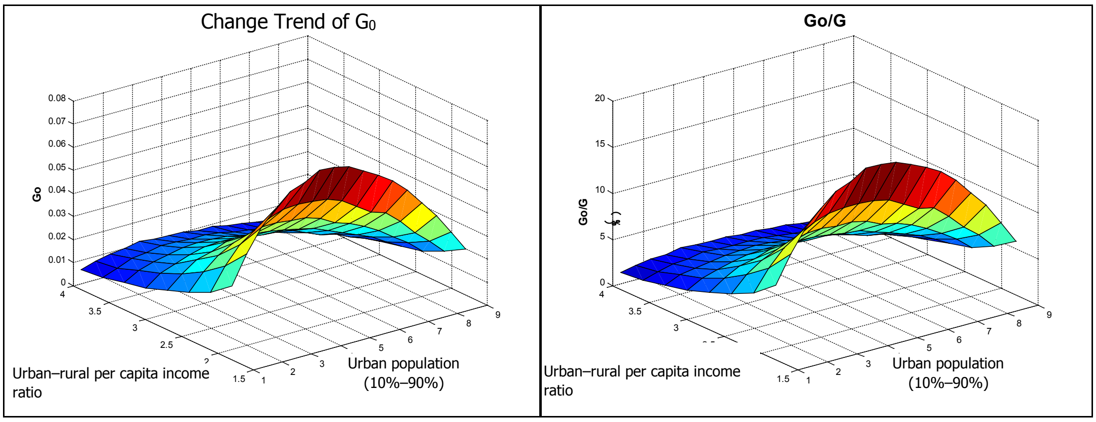
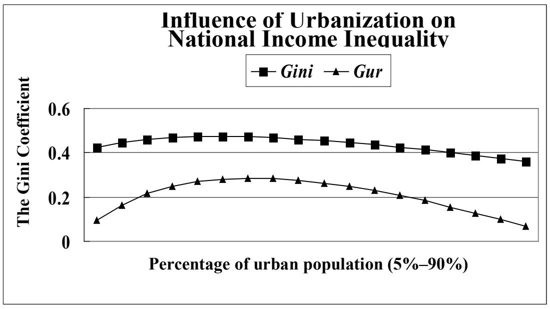


Figure 3. Change trend for G_0



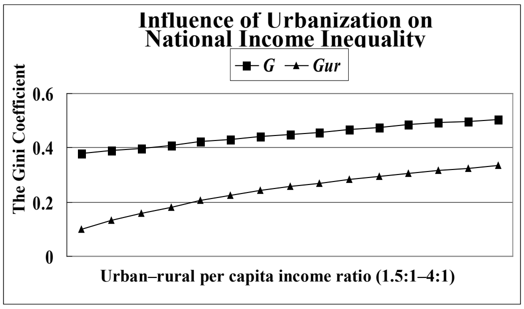


Figure 4. Urbanisation process and urban–rural income gap

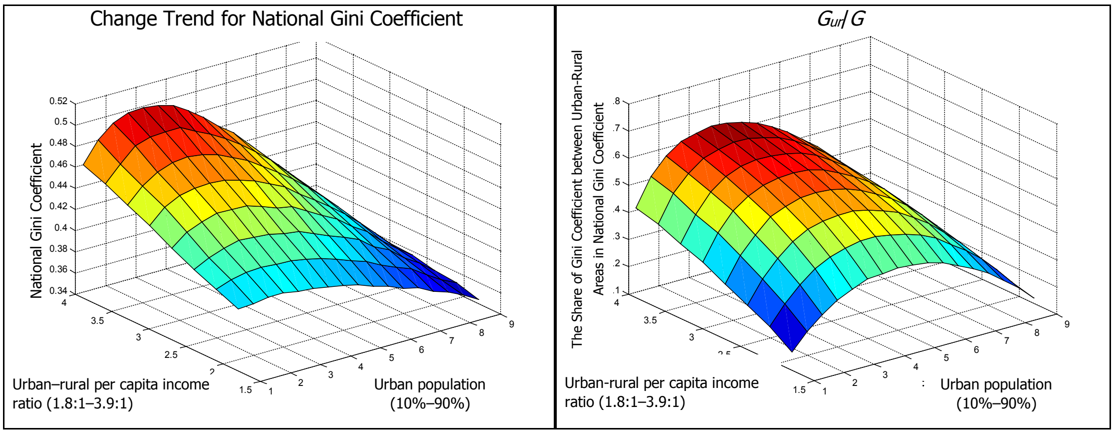


Figure 5. National Gini index and G_{ur}/G trends