On Income Inequality within China's Provinces

Prabir Bhattacharya, Javier Palacio-Torralba, Xinrong Li

Heriot Watt University, Riccarton, Edinburgh, UK
Email: P.C.Bhattacharya@hw.ac.uk

Abstract

The paper provides estimates of Gini coefficients for each of the 31 Chinese provinces for the years 2000-2012. The estimates bring out the extremely high income inequality to be found in nearly half of the provinces in sharp contrast to the extremely low income inequality to be found in the other half. The country seems to be sharply divided into two extremes when it comes to considering the extent of income inequality within its provinces. Policy implications of the findings are considered.

Keywords

Income Inequality, Gini Coefficients, Chinese Provinces, Support Vector Machines, Policy Implications

1. Introduction

This paper makes an attempt at providing Gini coefficient estimates for each of the 31 Chinese provinces for the years 2000 to 2012. We also make some observations on the relationships between the estimated Gini coefficients and the average years of schooling (a proxy for human capital), the physical capital stock, and the rural and urban incomes.

As is well known, Gini coefficient is the most widely used measure of income inequality. It ranges from 0 (perfect equality) to 1 (perfect inequality), with a higher value of the Gini coefficient suggesting a higher inequality of income distribution. For China as a whole, a number of different estimates of the Gini coefficients are available from different sources.[1] The official estimates of the Gini coefficients for the years 2003 to 2012 were provided by the Director of the National Bureau of Statistics of China (NBS) at a press conference in January 2013.[2] These official estimates are presented in Table 1. Estimates for the years 2000,

---

[1]See, among others, Chen and Hou (2008), Deininger and Squire (1996), Li et al. (2013), Tian (2012), and Xie and Zhou (2014).
[2]See Gan (2014) and Xie et al. (2013).

Table 1. Gini coefficient estimates for China: 2000-2012.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Gini	0.3951	0.405	0.4253	0.479	0.473	0.485	0.487	0.484	0.491	0.49	0.481	0.477	0.474

Source: National Bureau of Statistics of China for 2003-2012; Tian (2012) for 2000-2002.

2001 and 2002 reported in Table 1 are those from Tian (2012).

However, while there are a number of estimates of the Gini coefficient available at the national level, Gini coefficient estimates for each of the Chinese provinces are scarcer. Apart from its curiosity value, for policy purposes, too, it would appear to be of some importance to know the extent of income inequality within each of the provinces. Thus, the case for redistributive policies, ceteris paribus, would be stronger in provinces with higher income inequality than those with lower income inequality. Further, in designing, say, the most appropriate social and health policies, the degree of income inequality within a province would clearly be an important variable to consider. Indeed, even for the designing of national macroeconomic policies, the policy makers may find it useful to factor in the degree of income inequality within each province in their decision making process. For these and other reasons (including political and even ecological ones), computations of the Gini coefficients at the provincial level would appear to be of some importance.

2. Data, Method and Gini Coefficient Estimates

The data for our computations of the provincial Gini coefficients come from the China Statistical Yearbook of various years. We have data on rural and urban populations for each of the 31 provinces. The data on per capita disposable income of rural and urban residents are available from the Yearbooks as are the data on GDP per capita and average years of schooling.[3] The data on exact income distribution in each of the provinces is not available. If we assume a uniform income distribution (i.e., assume that every person in a given province has the same income, which in any case would be untrue), the estimated Gini coefficients would obviously be much lower than those reported in Table 1, besides being greatly misleading. We, therefore, need a different tack.

Organizing the data on a histogram it is seen that both the urban and rural income of the country approximately follow a lognormal distribution (see Figure 1). We, therefore, assume that the urban and rural incomes in each of the provinces also follow a lognormal distribution. In that case, Gini coefficient can

[3]Average years of schooling is measured based on the current educational system in China: 16 years for graduates from universities; 12 years for high school graduates; 9 years for middle school graduates; 6 years for primary school graduates and zero years for non-educated persons. Since children usually start their education at the age of 7 in China, the population size is based on the people who are over the age of 6. The data set can be found in the China Statistical Yearbook.

The provincial capital stock is estimated following Shan (2008). See also Wang (2011). For per capita GDP, see also CEINET (China Economy Information NET).
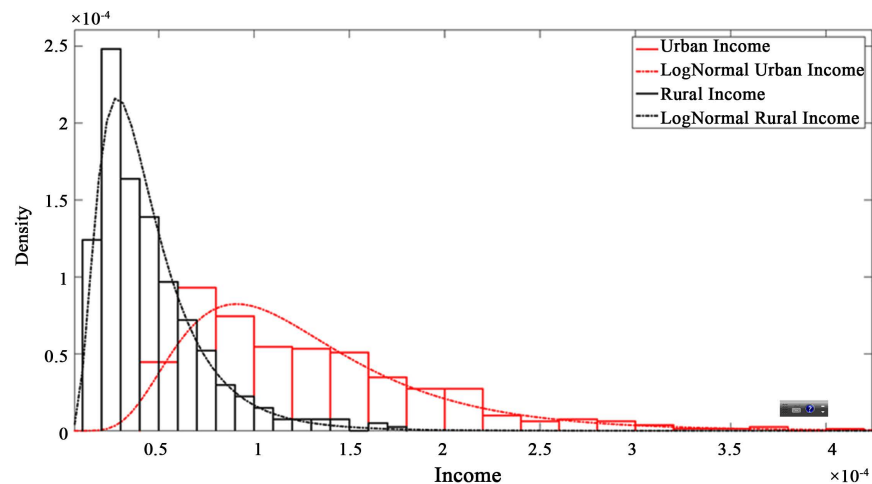
Figure 1. Distribution of the urban and rural income in the 31 provinces of China during the period 2000-2012. The lognormal fitting (dashed lines) match the binned data.

be calculated asas erf(var)[4]. The problem consists then of finding the values of the variance for each province so that the difference between the estimates in **Table 1** and our estimates is minimized in a least squares sense. This leads to a 31 variable nonlinear minimization problem solved using a gradient type algorithm combined with a multistart approach to try avoiding local minima. We constrain the minimum variance to be 0.05 which leads to a Gini coefficient of 0.056 to avoid the stagnation of the algorithm around 0, which would be an unrealistic solution. The resulting estimates of the Gini coefficients for the whole country are then seen to be very close to those reported in **Table 1** (see **Figure 2**). Using this approach, we then have the estimated Gini coefficients for each of the 31 provinces and these are presented in **Table 2**.

It will be noted that in **Table 2** we report the Gini coefficient estimates for only one year, the year 2012. This is because our computations of the Gini coefficients follow the assumption that the variations in the Gini coefficients over the whole period of the study would be low (so that the value of the Gini coefficient in any one of these years would not be that different from those of the other years).[5] Indeed, if we look at the estimates of the Gini coefficients for China as a whole reported in **Table 1**, it can be seen that these too remained fairly constant during this period (suggesting no significant changes in the distribution of income during the period).

What is particularly interesting about our estimated Gini coefficients is the wide variations in these coefficients across China's 31 provinces. Our estimates suggest that while 15 provinces have what may be considered highly equal

---

[4]erf denotes the error function defined as: $\mathrm{erf}(x) = \dfrac{2}{\mathrm{sqrt}(\pi)} \int_0^x \mathrm{e}^{-t^2} \mathrm{d}t$.

[5]Although it would be possible to consider a different variance for each province and year, doing so would lead to an ill-posed problem where many different minima could be found, thereby reducing the significance of any solutions obtained.

176

Table 2. Estimates of provincial Gini coefficients.

Province	Gini Coefficient	Province	Gini Coefficient
Beijing	0.97	Henan	0.06
Tianjin	0.85	Hubei	0.06
Hebei	0.79	Hunan	0.06
Shanxi	0.71	Guangdong	0.99
Inner Mongolia	0.26	Guangxi	0.06
Liaoning	0.20	Hainan	0.65
Jilin	0.06	Chongqing	0.06
Heilongjiang	0.06	Sichuan	0.06
Shanghai	0.98	Guizhou	0.06
Jiangsu	0.92	Yunnan	0.90
Zhejiang	0.97	Tibet	0.59
Anhui	0.06	Shaanxi	0.06
Fujian	0.84	Gansu	0.06
Jiangxi	0.84	Qinghai	0.56
Shandong	0.83	Ningxia	0.68
		Xinjiang	0.93
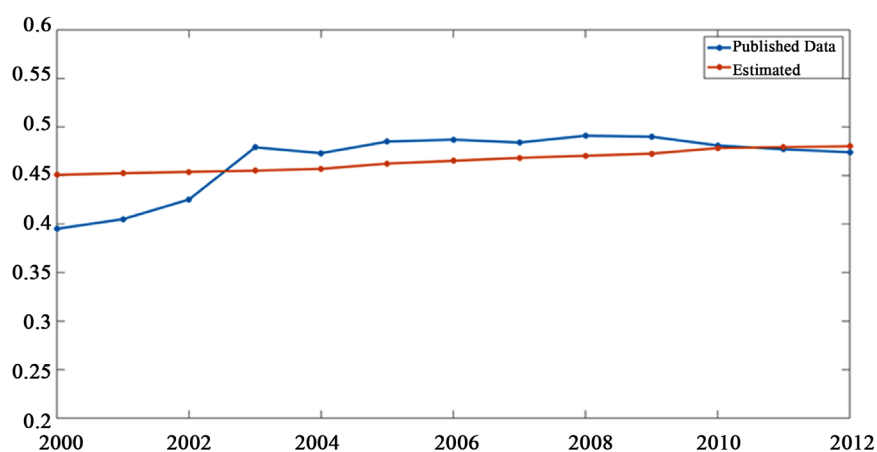
Figure 2. Comparison of our estimated Gini coefficients for China and those reported in Table 1.

income distribution (13 of these provinces have Gini coefficients of only around 0.06 and the other two, Lianoing and Inner Mongolia, have Gini coefficients of 0.197 and 0.264 respectively), the rest have a very high or extremely high inequality in the distribution of income (4 provinces have Gini coefficients between 0.56 and 0.68, and the remaining 12 provinces have Gini coefficients between 0.71 and 0.99). Guandong has an estimated Gini coefficient of 0.99, with Shanghai, Zhilang and Beijing not far behind with Gini coefficients of about 0.97. Our estimatedGini coefficients thus cluster, as it were, in two extremes. It is indeed

rather surprising to find even considering data limitations such extremes in the distribution of income within the provinces in the same country.

3. Gini Coefficients, Support Vector Machines and Results

As already mentioned, in addition to the data on urban and rural income and per capita GDP, we also have data on the average years of schooling and the physical capital stock. It may, therefore, be of some interest to see what relationships, if any, we can find between our estimated Gini coefficients and these variables of interest during the period of our study. The findings of such an exercise may also, of course, be of some interest to the policy makers in assessing the consequences of any particular patterns of growth on income distribution as well as assessing the consequences of any redistributive policies.

To explore these relationships, we use support vector machines (SVM).[6] To validate the models a leave-one-out cross validation approach is used. In what we call the Type 1 Gini coefficient classification, we divide the estimated Gini coefficients into three classes: class 1 consisting of Gini coefficients lower than 0.21 (the low Gini coefficient class); class 2 consisting of Gini coefficients between 0.21 and 0.28 (the medium Gini coefficient class; in our case this class contains only one instance); and class 3 consisting of Gini coefficients larger than 0.28 (the high Gini coefficient class). Additionally, we also consider a second classificatory scheme Type 2 Gini coefficient classification with the bounds of 0.56 and 0.68 for the low and medium Gini coefficient classes respectively. A summary is presented in Tables 3 while Table 4 shows the predictions of the Gini coefficient classes vis-a-vis the variables of interest.

The results of the SVM analysis are presented in Figure 3 and Figure 4. Red

Table 3. Summary of Type 1 and Type 2 Gini coefficient classifications.

	Gini (Type 1 classification)	Gini (Type 2 classification)
Low (Class 1)	Gini < 0.2	Gini < 0.56
Medium (Class 2)	$0.2 \leq$ Gini < 0.28	$0.56 \leq$ Gini < 0.68
High (Class 3)	$0.28 \leq$ Gini ≤ 1	$0.68 \leq$ Gini ≤ 1

Table 4. Prediction of the Gini coefficient classes with support vector machines for the variables considered.

	Gini (Type 1 classification)	Gini (Type 2 classification)
Average Years of Schooling	63.05%	60.7%
Physical Capital Stock	58.06%	68.3%
Rural Income	58.65%	61.29%
Urban Income	60.12%	59.94%

[6]SVM is a popular class of learning algorithms widely used in classification and regression analysis. The method consists of finding a hyperplane or hypersurface that maximizes the distance between elements that belong to different classes.
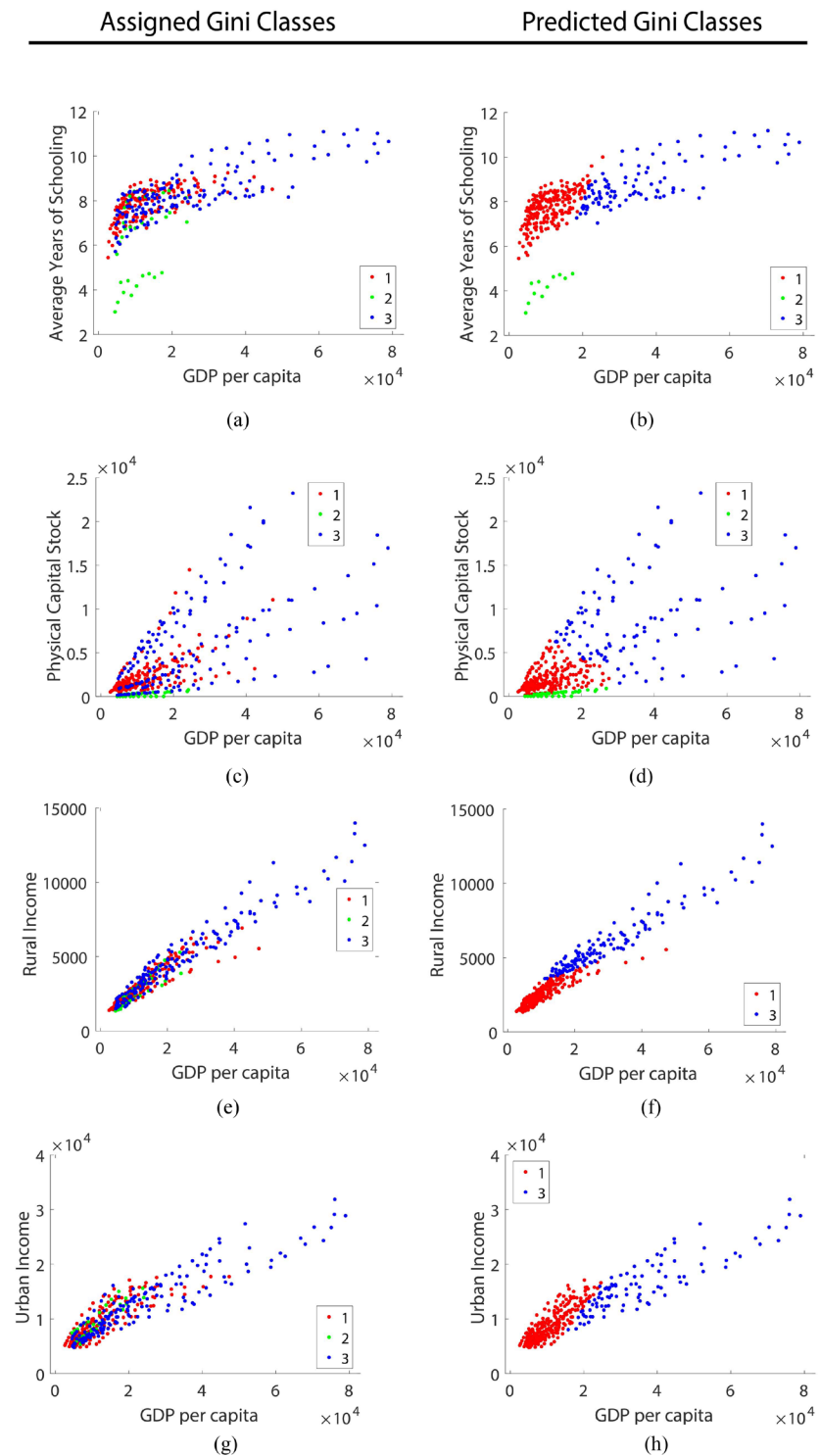
Assigned Gini Classes   Predicted Gini Classes

Figure 3. Assigned and predicted Gini classes for Type 1 Gini coefficient classification.

dots denote class 1 Gini coefficients, green dots class 2 and blue dots class 3 Gini coefficients. **Figure 3(a)** and **Figure 3(b)** show the distribution of the assigned Gini coefficient classes for the average years of schooling under the Type 1 classification. The SVM separates only class1 and class 3 due to the low number of
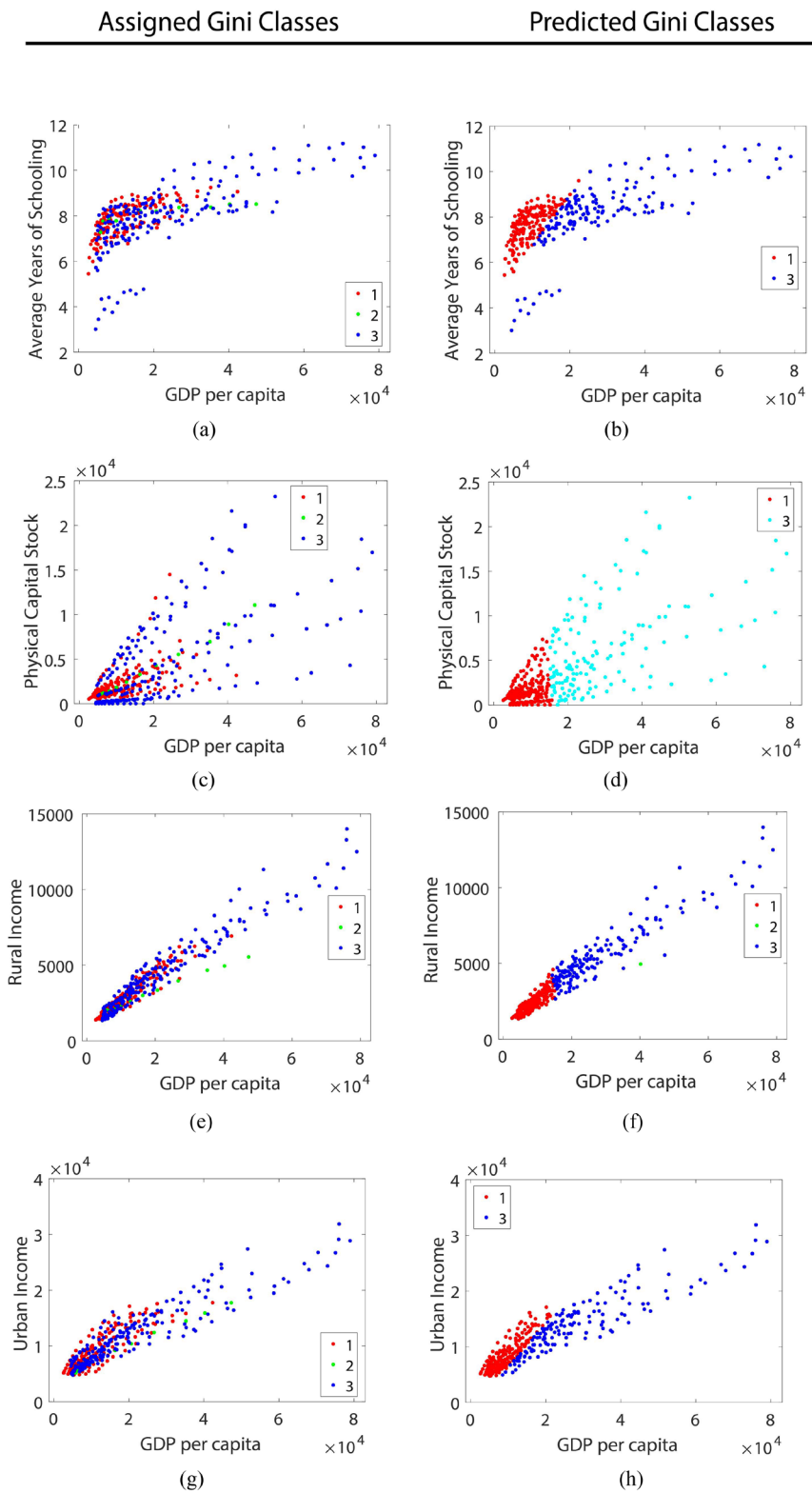
Assigned Gini Classes   Predicted Gini Classes

Figure 4. Assigned and predicted Gini classes for Type 2 Gini coefficient classification.

samples in class 2 (only one case in class 2). As can be seen, low Gini coefficients are expected only for low GDP per capita and medium level of years of schooling

(around 6 - 9 years). The SVM traces a line that differentiates the two classes with an accuracy of 63.05%.

The results for the physical capital stock (Figure 3(c) and Figure 3(d)) are seen to be more dispersed. Most of the low Gini coefficient cases are found for low GDP per capita (Figure 3(c)). The SVM traces a vertical decision line for GDP per capita of around 17,500 Yuan (Figure 3(d)), with a classification accuracy of 58.06%. At a GDP per capita of higher than 17,500 Yuan, we have high Gini coefficient cases, but the physical capital stock can be both high and low. The amount of physical capital stock, in other words, does not seem to make much of a difference as to whether the Gini coefficient will be high or low. This, it will be noted, is in contrast to the human capital (proxied by the average years of schooling) case where, as we have just seen, high Gini coefficient cases are expected only for very high average years of schooling.

Turning to the results for the rural and urban incomes, Figure 3(e) shows the distribution of the assigned Gini coefficient classes for the rural income. Compared to the case of the urban income (Figure 3(g)), the boundaries between the upper and lower income for a given GDP per capita is smaller in the case of the rural income. It would appear that for a given per capita GDP, inequality is more widely dispersed in the urban than in rural areas. Also, the actual level of per capita GDP does not seem to make a great deal of difference to the extent of inequality in rural areas.[7]

The results discussed above are for the Type 1 classification scheme. The results for the Type 2 classification are presented in Figure 4. While there are similarities, there are also some differences in the results under the Type 2 classification compared to those under the Type 1 classification scheme. When considering the per capita GDP, the SVM in the Type 2 classification case fails to correctly predict class 2 (medium Gini coefficient) provinces which are somewhat randomly mixed with the other two classes. The separation between class 1 and class 2 occurs at around 20,000 Yuan with a lower slope than in the Type 1 classification. This would mean that the algorithm finds some relation around the 17,000-20,000 Yuan to separate the high and low Gini coefficient classes for a large range of bounds for high and low Ginicoefficient classes. Also, vis-à-vis the rural income, the decision line is quasi horizontal in the Type 2 classification case. While this might appear as a significant difference, in practice, however, only a handful of instances are affected.

4. Conclusion

In this paper we have presented Gini coefficient estimates for the Chinese provinces for the period 2000 to 2012. We have also made some observations on the relationship between the estimated Gini coefficients and the average years of schooling (a proxy for human capital), the physical capital stock, and the rural

---

[7]It is important to notice the difference in the slope of the resulting decision line in both cases. It is much steeper in the rural case, therefore containing the low Gini coefficients for lower values of GDP per capita.

and urban incomes. In the absence of the availability of the exact income distribution data for each of the 31 provinces, our estimates presented here are of a preliminary nature. Nevertheless, these estimates, we believe, do give some idea about the evolution of income distribution within these provinces during the period of the study (the main assumption here being that income distribution probably did not change a great deal during this period) as well as showing the extremely high inequality to be found in nearly half of the provinces, in sharp contrast to the extremely low income inequality to be found in the other half. The country seems to be sharply divided into two extremes when it comes to considering the extent of income inequality within its provinces. This is a result we did not anticipate and one that we find particularly interesting. This suggests that there is room for the overall Gini coefficient to decline in China if growth can be shifted towards the provinces with low income inequality. Similarly, the SVM analysis suggests that any redirection of growth towards the rural areas would also improve income distribution in the country.

References

Chen, J., & Hou, W. (2008). New Approach to Estimate the Chinese Gini Coefficients from 1978-2003. *19th Annual Conference of the Chinese Economic Association (CEA)*.

Deininger, K., & Squire, L. (1996). A New Data Set Measuring Income Inequality. *The World Bank Economic Review, 10,* 565-591. https://doi.org/10.1093/wber/10.3.565

Gan, L. (2014). *Data You Need to Know About China.* Berlin: Springer. https://doi.org/10.1007/978-3-642-38151-5

Li, S., Sato, H., & Sicular, T. (2013). *Rising Inequality in China: Challenges to a Harmonious Society.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139035057

National Bureau of Statistics of China. China Statistical Yearbook, Various Years.

Shan, H. J. (2008). Reestimating the Capital Stock of China: 1952-2006. *The Journal of Quantitative & Technical Economics, 10,* 17-31. (In Chinese)

Tian, W. (2012). The Chinese Gini Coefficient Calculation and Its Trend Analysis. *The Journal of Humanities, 2,* 56-61. (In Chinese)

Wang, C. (2011). A Study of the Dynamic Relationship between Physical Capital, Human Capital and Economic Growth. *Journal of Shandong University of Finance, 5,* 80-89.

Xie, Y., & Zhou, X. (2014). Income Inequality in Today's China. *Proceedings of the National Academy of Sciences, 111,* 6928-6933. https://doi.org/10.1073/pnas.1403158111

Xie, Y., Zhang, X., Xu, Q., & Zhang, C. (2013). In Y. Xie, et al., (eds), *Wellbeing Development Reports of China* (pp. 27-53). Beijing: Peking University Press. (In Chinese)