DEPARTMENT OF ECONOMICS

ISSN 1441-5429

DISCUSSION PAPER 32/15

Income inequality in China: Testing the Kuznets Hypothesis with National Time Series and Provincial Panel Data 1978-2011*

Wenli Cheng[a] and Yongzheng Wu[b]**

May 2015

Abstract:

This paper investigates income inequality in the post-reform Chinese economy using both national time series and provincial panel data 1978 to 2011. We identify a Kuznets inverted-U relationship between economic development and income inequality and show that this relationship was driven by the process of urbanization. We find that, after controlling for urbanization, low productivity in agriculture relative to that of the economy as a whole (i.e., dualism) and inflation appear to have been significant contributing factors to income inequality. There is also some evidence to suggest that, the expansion of higher education may have widened income inequality, but the expansion of secondary education may have narrowed it.

Keywords: Kuznets curve, income inequality in China, Theil index, urbanisation, dualism

JEL Classification Numbers: O15, O53

* This paper is an extension of our earlier paper entitled "Understanding the Kuznets Process: An Empirical Investigation of Income Inequality in China 1978-2011", Discussion paper 12-14, Department of Economics, Monash University. The earlier paper examines only the national time series data.

**The project was supported by China Postdoctoral Science Foundation (Grant No. 2012M510736) and by College of Mathematics and Computer Science, Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China), Hunan Normal University, Changsha, Hunan 410081, P. R. China.

[a] Department of Economics, Monash University, Australia

[b] China Economics and Management Academy, Central University of Finance and Economics and College of Mathematics and Computer Science, Hunan Normal University

Income inequality in China: Testing the Kuznets Hypothesis with National Time Series and Provincial Panel Data 1978-2011

1. Introduction

The Chinese economy has experienced phenomenal growth since 1978 when its transition to a market economy began. Between 1978 and 2011, real GDP per capita (at constant 2005 prices) grew from 1,582 to 27,309 Chinese yuan (CNY), which amounts to an average annual growth rate of 9.15 %. Initially, the economic growth also reduced income inequality; however income inequality rose substantially from mid 1980s to mid 1990s. The second half of 1990s saw some reduction in inequality, but it did not last. Inequality continued to rise, reaching a peak in mid 2000s before showing some signs of improvement (see Figure 1). The broad pattern of income inequality gives rise to three questions: (1) is rising income inequality is an inevitable "side effect" of early stage economic development? (2) Can we expect inequality to fall as the economy develops further? Or in other words, is there a Kuznets inverted-U relationship between income inequality and economic development in China? (3) What are some of the contributing factors to income inequality that we might be able to influence through policy?

This paper uses both national time series and provincial panel data to investigate whether there was a Kuznets relationship between economic development and income inequality in China during the post reform period of 1978-2011. It also studies other factors that may contribute to the observed inequality pattern. The main findings of our analysis are: (1) there was a Kuznets relationship between economic development and income inequality; (2) a driving force behind the non-linearity of the Kuznets process was urbanization; (3) after controlling for urbanization, low productivity in agriculture relative to that of the economy as a whole (i.e., dualism) and inflation appear to have been significant contributing factors to income inequality; (4) the expansion of higher education may have widened income inequality, but the expansion of secondary education may have narrowed it. However the effect of education on inequality do not seem to be robust.

The rest of the paper is organized as follows. Section 2 reviews the related literature and explains how this paper contributes to it. Section 3 presents the empirical model and describes the data used in this study. Section 5 analyzes the estimation results. Section 5 concludes with some policy implications.

2. Literature review

This paper belongs to the broad literature on the relationship between economic development and income inequality. Based on the statistical regularities he observed from historical economic data of England, Germany and the United States, Kuznets (1955) suggests that there is an inverted-U relationship between inequality and development: with inequality "widening in the early phases of economic growth when the transition from the pre-industrial to the industrial civilization was most rapid; becoming stabilized for a while; and then narrowing in the later phases" (p.18). This is the well-known Kuznets hypothesis.

In his original work, Kuznets (1955) emphasized two drivers behind his hypothesis: the concentration of savings and urbanization. As a rule, upper-income earners save more. The cumulative effect of this savings concentration is that an increasing proportion of assets would be held by upper-income earners thereby increasing their income share. However, there factors that counteract this savings concentration, for example, income redistribution policies, the increasing importance of service income, and the dynamism of a growing economy that offers more individual opportunities.

On the role of urbanization, Kuznets (1955) contends that income tends to be more unevenly distributed in urban areas, and that the income gap between urban and rural residents does not necessarily narrow with economic development.[1] Given these tendencies, urbanization raises the share of the more unequal of the two component distributions, which increases overall inequality. During later stages of development, the widening of overall income inequality associated urbanization is more than offset by the narrowing of inequality within the urban sector as new migrants better adapt to urban life

[1] Greater income disparity in urban areas may be due to greater occupational diversity and the large income gap between established professionals and recently arrived migrants.

and obtain greater political power to support their claims for a larger income share. Thus the income inequality path takes the shape of an inverted-U.

While the features of the urbanization process as described by Kuznets (1955) would explain an inverted-U relationship between inequality and development, other researchers have shown that the simple fact that urbanization enables some initially poorer rural individuals to earn a higher income in urban areas could explain the Kuznets hypothesis. Using a simple two-sector model, Robinson (1976) demonstrates that even if the mean income and the income distribution for the urban and the rural sector remain unchanged, the overall inequality (as measured by the log variance of income) is a quadratic function of the urban population share. In other words, in the two-sector economy, overall inequality will first rises and then falls as the share of urban population increases. Knight (1976) and Fields (1979) have obtained similar results with different measures of inequality. Knight (1976) explains the logic behind the inverted-U curve in the context of urbanization as follows. If everyone is initially in the rural sector and has the same low income, the Gini coefficient (G) is zero. If one person moves to the urban sector and receives a higher income without changing anyone else's income, G goes up slightly. As more people moves to the higher income sector, G continues to rise. When the number of people remaining in the lower-income rural sector falls to a certain level, G starts to fall. Therefore the process of urbanization would be accompanied by an initial increase and a subsequent decline of overall measured inequality.

A number of early cross-country empirical studies have confirmed the Kuznets relationship between income inequality and development (see for instance, Ahluwalia, 1976, Lecaillon et al., 1984). However these studies have been criticized on both methodological and data comparability grounds (Saith, 1983; Adelman and Robinson, 1989, Anand and Kanbur 1993a). It is argued that inter-temporal national studies rather than cross-country analyses are required to test the Kuznets hypothesis (Saith, 1983). As an empirical investigation of the relationship between development and inequality in China over the post-reform period of 1978-2011, our paper provides a useful test of the Kuznets hypothesis. To our knowledge, few studies have specifically tested the Kuznets hypothesis in the Chinese context. One exception we find is Zhang et al. (2012) who, in the process of examining the effects of financial development on urban-rural inequality in

China over the period 1978-2006, also identified an inverted-U relationship between urban-rural income gap and per capita real GDP. Different from Zhang et al. (2012), we focus on urbanization as the driver behind the Kuznets relationship in line with Kuznets' original conjecture and subsequent theoretical work discussed above. Moreover, we consider a longer time period from 1978 to 2011 and use both national time series and provincial panel data.

Our paper is related to the large literature on the pattern and determinants of income inequality in China (see, for example, Kanbur and Zhang, 1999, 2005; Ravallion and Chen, 2007; Sicular et al. 2007; Zhang and Zou, 2012). A key finding of this literature is that rural-urban income disparity accounts for a large share of overall inequality in China. Different studies have focused on different factors that affect income inequality. Factors that have been found to increase income inequality in China include urban-biased government expenditure and investment (Yang, 1999, 2002; Lin and Chen, 2011; Zhang and Zou 2012); the *houkou* (i.e., household registration) system that restricts labor mobility (Lin et al. 2004; Fan et al. 2011); high tax burden on rural residents (Tao et al., 2004); and inflation (Ravallion and Chen, 2007). Zhang et al. (2012) has also found secondary education to be positively associated with urban-rural inequality. They suggest that this is due to the fact that education attainment is higher in urban areas.

Apart from testing urbanization as the driver behind the Kuznets inverted-U relationship between income inequality and development, we pay special attention to the role of dualism on inequality. Dualism indicates the degree to which agricultural productivity lags behind the productivity of the economy as a whole. Nielsen (1994) and Bourguignon and Morrisson (1998) have found dualism to be an important explanatory factor of income inequality in cross-country studies, however, it has been largely neglected in studies of inequality in China. We also investigate the effects of inflation and education on inequality. In particular, we look into whether higher education and secondary education may have different effects on inequality in China, which is a question not addressed in the existing literature.

3. Model specifications and data

We consider four factors that determine income inequality (*TT*) in China: urbanization (*URBAN*), dualism (*DUAL*), inflation (*INF*), and education (*EDU*)

$$TT = f(URBAN, DUAL, INF, EDU) \quad (1)$$

Where *EDU* may be either higher education (*HEDU*) or secondary education (*SEDU*).

Based on equation (1), we can specify two empirical models, one with higher education and another with secondary education:

$$\ln(TT) = \alpha_0 + \alpha_1 \ln(URBAN) + \alpha_2 (\ln(URBAN))^2 + \alpha_3 \ln(DUAL) + \alpha_4 \ln(INF) + \alpha_5 ln(HEDU) + \alpha_6 T1992 + u \quad (2a)$$

$$\ln(TT) = \beta_0 + \beta_1 \ln(URBAN) + \beta_2 (\ln(URBAN))^2 + \beta_3 \ln(DUAL) + \beta_4 \ln(INF) + \beta_5 ln(SEDU) + \beta_6 T1992 + \varepsilon \quad (2b)$$

where *T*1992 is time dummy variable which equals to 0 for years 1978-1992, and 1 otherwise.

Equations (2a) and (2b) specify a non-linear relationship between urbanization and inequality in line with the Kuznets hypothesis. As noted in the last section, the driving force behind the non-linear relationship between income inequality and development may be the urbanization process. That is, as an economy develops, a larger share of the population moves to urban areas and earn a higher income. This movement leads to an initial increase and a subsequent fall in inequality (Kuznets, 1995; Knight, 1976). If the Kuznets relationship applies to the Chinese experience, we would see the coefficients of ln(*URBAN*) (i.e., α_1 and β_1) to be positive and the those of $(\ln(URBAN))^2$ (i.e., α_2 and β_2) to be negative.

The second determinant of income inequality in our model is dualism (*DUAL*). As noted earlier, dualism is a measure of productivity difference between agriculture and the rest of the economy. Standard neoclassical economic theory postulates that if marginal productivity is higher in one sector than another, factors of production would be attracted to the sector with higher marginal productivity. Factor movement would continue until marginal productivities in all sectors are equalized, which means factor income should also tend to equalize. In real economies, however, such factor movements may be significantly constrained so that dualism results which in turn produces income disparity across sectors. In China, labor movements are restricted by the "hukou" system of household registration, and capital allocation is also biased in favor of the urban sector, both leading to dualism. Dualism affects inequality because productivity differences correspond to different income-generating abilities. The higher the degree of dualism, that is, the more productivity in agriculture lags behind that in other sectors, the lower income rural residents are likely to earn relative to urban residents. Thus we expect the coefficients of $\ln(DUAL)$ (i.e., α_1 and β_1) to be positive.

The third determinant of inequality in our model is inflation (*INF*). The study of re-distributional effect of inflation can be traced back to Cantillon (1755), who links inflation to an increase in money supply. He contends that where there is an increase in money supply, the new money enters the economy *at a specific point*, which means some people receive the new money first. The first receivers of new money spend it, so the money reaches their suppliers who in turn pass it on through their own purchases. In this way, the new money permeates the economy via multiple *sequential* transactions. The early recipients of the new money benefit at the expense of the late receivers because the former see their income increase before prices increase for all the goods they buy; whereas the late recipients experience higher prices before their income levels rise. Since higher income earners tend to be politically more powerful and have better access to finance, they are more likely to receive the new money first and benefit from inflation (Bai and Cheng, 2014). That is, inflation driven by a monetary expansion would redistribute wealth from the poor to the rich, thereby exacerbating inequality.

On the other hand, Lewis (1954) argues that in a dual economy with "unlimited supplies of labor", credit creation can facilitate the employment of more labor to speed up capital formation. The expansion of credit will lead to inflation in the short run, but prices will fall once more output is produced as more capital is put to use. Before more output is produced however, the existing quantity of output is redistributed to the newly employed workers at the expense of the rest of the community and the income share of capital owners rises as more capital is accumulated. The increased employment tends to reduce income inequality but the higher share of capital income tends to raise it, so the net effect depends on the relative magnitudes of the two forces.

While the theories do not give a clear prediction about inflation's net effects on inequality, we suspect that inflation driven by credit expansions in China had more of the effect of enriching the privileged class than creating job opportunities benefiting the poor. Thus we hypothesize that inflation had a net effect of widening inequality in China, that is, we expect α_4 and β_4 to be positive.

The fourth determinant of inequality in our model is education. It is generally believed that in the long run, education is an important income equalizer for at least two reasons. First, low income families can more easily acquire human capital through education than accumulate physical or financial capital through savings or inheritance. Secondly, unlike physical capital accumulation that is prone to concentration, the expansion of human capital involves dispersion of knowledge and skills across the wider population (Ahluwalis, 1976). However, in the short run, education expansion may be associated with higher inequality. For instance, if people from high income families have better education opportunities, overall inequality may increase during the course of education expansion (Nielsen, 1994). Also, the income gap between the educated and the uneducated may increase as skill-biased technological change in recent decades has raised the return to education (Acemoglu, 2002). Moreover, in the Chinese context, as migrants to urban areas tend to be more highly educated, the brain drain in rural areas hinders rural sector productivity growth, thereby aggravating urban-rural income inequality.

In the existing literature, education expansion is typically measured by secondary school enrollment. In this paper, we measure education separately by higher education enrollment and secondary school enrollment, and examine whether they had different effects on inequality. To the extent that higher education is one path for talented young people in rural areas to find highly-paid employment in cities, the expansion of higher education may result in brain drain in rural areas, thereby widening rural-urban inequality. Secondary education expansion on the other hand may have a different effect. As an important way of accumulating human capital, secondary education improves the labor productivity and income earning abilities of all those receiving the education. The expansion of secondary education is likely to benefit the rural region more because the rural region started from a lower secondary school enrollment rate, and would receive a relatively greater improvement in education opportunities. Thus, we hypothesize that higher education expansion would have a inequality-widening effect, whereas secondary education expansion would have an inequality-narrowing effect. That is, we expect α_5 to be positive and β_5 to be positive.

We also include a time dummy (*T1992*) in our empirical model, which imposes a structural break at the end of year 1992. In choosing the timing of the structural break, we took into account two main factors. First, China's adoption of the United Nations System of National Accounts 1993 marked a major step towards an international standard of national accounting. Second, following Deng Xiaoping's southern tour in 1992, the Chinese central government endorsed the notion of "socialist market economy" and the sped up the pace of economic reforms.

We use both national time series and provincial panel data for the period 1978-2011 to estimate equations (2a) and (2b). The provincial panel data contain information for 31 province-level divisions of administrative areas (which includes 22 provinces, 5 autonomous regions and 4 directly-administered municipalities). The time series and panel data for 1978-2008 are from *China Compendium of Statistics 1949-2009*. The time series data for 2009-2011 are from 2010-2012 issues of *China Statistical Yearbook*. The panel data for 2009-2011 are from 2010-2012 issues of *China Statistical Yearbook for Regional Economy*.

The definitions of all variables in our model together with their corresponding data sources are presented in Table 1. We provide further details below.[2]

The Theil index (*TT*) is our measure of income inequality. We have computed *TT* from provincial data on rural and urban incomes and populations (see the Appendix 1 for calculation details).

URBAN is the degree of urbanization measured by the share of urban population in total population. The degree of urbanization has increased substantially over our data period. In 1978, about 17.9% of the population resided in urban areas. By 2011, the figure had risen to 51.3%.

[2] The national time series data are in Appendix 2. Panel data can be obtained from the corresponding author.

DUAL is measured by the inverse of agricultural labor productivity relative to labor productivity for the economy as a whole, so that a larger value of *DUAL* indicates a lower relative productivity in agriculture. Since the primary sector in China contains mainly agriculture, it is often treated as being "equivalent to" agriculture in the literature (Fan et al. 2003). We thus use primary sector productivity as a proxy for agricultural productivity. *DUAL* fell from 2.5 in 1978 to 1.99 in 1984; then started to rise, reaching a peak of 3.8 in 2003. In 2011, *DUAL* remained at a high level of 3.5.

INF is measured by the consumer price index series with preceding year =100.

HEDU is higher education enrollment per 10,000 population (lagged by 5 years). Higher education enrollment increased substantially from 3.52 in 1973 to 132.28 in 2006.

SEDU is secondary education enrollment per 100 population (lagged by 5 years). Secondary enrollment increased from 3.86 in 1973 to 7.82 in 2006.

4. Estimation results

We conduct our time series estimation of equations (2a) and (2b) using the Autoregressive Distributed Lag model (ARDL) advocated by Pesaran (1997) and Pesaran and Smith (1998). This approach has been widely used in time series analyses, including studies of inequality (see, for instance, Jalil, 2012). The ARDL procedure consists of three steps. The first step involves selecting the appropriate lag orders of the ARDL model using either the Akaike Information Criterion (AIC) or the Schwartz Bayesian Criterion (SC). A variable Addition Test (ARDL case) is conducted to see whether there exists a long-run relationship among the variables. If the null hypothesis of no co-integration is rejected, one proceeds to the second step of estimating the long-run relationship using the selected ARDL model. In the third step, an error correction model is estimated, providing information on the speed of adjustment back to the long-run equilibrium following a shock.

We use Generalized Method of Moments (GMM) (Hansen, 1982) for our panel data estimation. GMM is widely considered to be an estimator well-suited to deal with potential endogeneity problems. In our estimation, explanatory variables lagged one period are used as instruments.

Before the models are estimated, we first test whether the variables under consideration are stationary. The test results for the time series and panel data are reported in Table 2a and Table 2b, respectively. The results suggest that all variables in first differences are stationary, which means that our estimation methods can be applied. For the time series, we also test the existence of a long term relationship among the variables (which is the second step of the ARDL method as described earlier). The test (reported in Table 3) indicates that a long run relationship exists for each model.

The results from the estimating our empirical model (equation (2a) and (2b)) are presented in Table 4. Estimation 1 and Estimation 2 are time series estimations with education measured by higher education enrollment (*ln(HEDU)*) and secondary school enrollment (*ln(SEDU)*), respectively. Estimation 3 and Estimation 4 are panel estimations with education measured by higher education enrollment (*ln(HEDU)*) and secondary school enrollment (*ln(SEDU)*), respectively.

As shown in Table 4, the coefficient of $\ln(URBAN)$ is positive and significant; and that of $LNURBAN^2$ is negative and significant in all 4 estimations. This is consistent with the theoretical prediction that urbanization is an important driver behind the Kuznets process. It indicates that the Chinese development experience confirms the Kuznets hypothesis that there is an inverted-U relationship between income inequality and development.

In all estimations except Estimation 2 (time series with education measured by secondary school enrollment), *ln(DUAL)* has a significant positive effect on *ln(TT)*. Noting that a high value of *ln(DUAL)* means low agricultural productivity relative to productivity of the economy as a whole, the positive coefficient of *ln(DUAL)* confirms our conjecture that low productivity in agriculture is likely to be associated with high overall income inequality. The importance of dualism in explaining inequality in China is consistent with the fact that a substantial proportion of overall inequality in China is attributable to urban-rural inequality (Lin and Chen, 2011). This result is also in line with the findings of Nielson (1994) and Bourguignon and Morrison (1998).

In all 4 estimations *ln(INF)* has a positive and significant effect on *ln(TT)*. This lends some support to our conjecture that the inflation in China benefited the rich and privileged (in the form of easier access to credit) more than poor (in the form of short term employment opportunities). Therefore the net effect of inflation on inequality was positive.

The coefficient of *Ln(UEDU)* is positive and significant in the time series estimation (Estimation 1), but is insignificant in the panel estimation (Estimation 3). In contrast, the coefficient of *ln(SEDU)* is insignificant in the time series estimation (Estimation 2), but negative and significant in panel estimation (Estimation 4). Thus we have some evidence to suggest that higher education expansion was associated with an increase in inequality. This is probably due to unequal education opportunities and brain drain from the rural sector. The expansion of secondary education seems to have the effect of narrowing inequality, which may be explained by the fact that rural areas benefited more from secondary school expansion because secondary education was already widespread in urban areas. The mixed results suggest that the effects of education are not robust. However since the sample size in the panel estimation is much larger than that in the time series estimation, the weight of the evidence seems to suggest that secondary education may be the preferred measure for education, and it appears that its expansion may have had the effect of narrowing inequality in China.

The time dummy variable *Y1992* is significantly positive (at 10% level) in Estimation 1, insignificant in Estimation 2, and significantly negative (at 1% level) in Estimations 3 and 4. That is, there is some weak time series evidence suggesting a lower level of inequality after 1992, and some (stronger) panel data evidence indicating a higher level of inequality after 1992.

5. Conclusion

In this paper, we have studied the pattern and determinants of overall income inequality in the post-reform Chinese economy of 1978-2011 using both national time series and provincial panel data. We have identified a Kuznets inverted-U relationship between income inequality and economic development and have shown that urbanization was an important driver of the Kuznets process. After controlling for urbanization, we have found that dualism and inflation appear to have been significant contributing factors to income inequality. We have also presented evidence which suggests, the expansion of higher education may have widened income inequality, but the expansion of secondary education may have narrowed it. However the effects of education on inequality do not seem to be robust.

A couple of implications following from the results of our paper are worth noting. First, since measured inequality rises with the increasing relative size of the higher-income urban population in the initial stages of development even if the relative average income between rural and urban residents remain constant (Knight, 1976; and Fields, 1979), measured inequality by itself does not give us sufficient information about the well-being of different social groups. To have a clear understanding of the welfare implications of inequality, it is important to also look at more detailed information instead of focusing on a single aggregate statistic. For instance, it will be informative to look at how population sizes change for groups of different income levels over time.

Secondly, the importance of dualism in explaining inequality (after controlling for urbanization) suggests that improving agricultural productivity not only enhances efficiency but also is likely to be an effective way of reducing inequality. From the beginning of the reforms in 1978 to the mid-1980s, agricultural productivity increased significantly with the implementation of the household responsibility system and with the rapid growth of township and village enterprises (TVEs) absorbing underemployed agricultural labor. During the same time, inequality fell substantially (see figure 1). The increased inequality in subsequent years may be partly attributable to urban-biased policies such as tightened state control of the financial sector severely hindering rural sector development (Huang, 2012). To address growing public concerns over inequality, policies should be directed to facilitate improvement in the rural sector. For instance, the rural sector's access to banking finance should be improved; the *hukou* system of household registration should be further relaxed to allow freer movement of labor between urban and rural areas; and the urban-bias in public investment spending should be corrected.

References

Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature, 40*(1), 7-72.

Adelman, I., & Robinson, S. (1989). Income distribution and development. In H. C. a. T. N. Srinivasan (Ed.), *Handbook of Development Economics* (Vol. II). Amsterdam: North Holland.

Ahluwalia, M. S. (1976). Inequality, poverty and development. *Journal of Development Economics, 3,* 307-342.

Anand, S., & Kanbur, S. M. R. (1993a). Inequality and development: A critique. *Journal of Development Economics, 41*(1), 19-43.

Bai, P. and Cheng, W. (2014). Who gets money first? Monetary expansion, ownership structure and wage inequality in China. Discussion paper, Economics Department, Monash University.

Bourguignon, F., and Morrisson, C. (1998). Inequality and development: the role of dualism. *Journal of Development Economics, 57*(2), 233-257.

Fan, S., Kanbur, R., & Zhang, X. (2011). China's regional disparities: experience and policy. *Review of Development Finance*(1), 47-56.

Fields, G. S. (1979). A Welfare Approach to Growth and Distribution in the Dual Economy. *Quarterly Journal of Economics, 93*(3), 325-353.

Hansen, P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 57, 1121-1152.

Huang, Y. (2012). How Did China Take Off? *Journal of Economic Perspectives, 26*(4), 147-170.

Jalil, A. (2012). Modeling Income Inequality and Openness in the Framework of Kuznets Curve: New Evidence from China. *Economic Modelling, 29*(2), 309-315.

Kanbur, R., & Zhang, X. (1999). Which Regional Inequality? The Evolution of Rural-Urban and Inland-Coastal Inequality in China from 1983 to 1995. *Journal of Comparative Economics, 27*(4), 686-701.

Kanbur, R., & Zhang, X. (2005). Fifty Years of Regional Inequality in China: a Journey Through Central Planning, Reform, and Openness. *Review of Development Economics, 9*(1), 87-106.

Knight, J. B. (1976). Explaining Income Distribution in Less Developed Countries: A Framework and an Agenda [Article]. *Oxford Bulletin of Economics & Statistics, 38*(3), 161-177.

Kuznets, S. (1955). Economic Growth and Income Inequality. *American Economic Review, 45*(1), 1-28.

Lecaillon, J., Paukert, F., Morrisson, C., & Germidis, D. (1984). *Income Distribution and Economic Development: An Analytical Survey*. Geneva, Switzerland: International Labor Office.

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School, 22*(2), 139-191.

Lin, J. Y., & Chen, B. (2011). Urbanization and Urban-Rural Inequality in China: A New Perspective from the Government's Development Strategy. *Frontiers of Economics in China, 6*(1), 1-21.

Lin, J. Y., Wang, G., & Zhao, Y. (2004). Regional Inequality and Labor Transfers in China. *Economic Development and Cultural Change, 52*(3), 587-603.

Nielsen, F. (1994). Income inequality and industrial development: Dualism revisited. *American Sociological Review, 59*(5), 654-677.

Pesaran, M. H. (1997). The Role of Economic Theory in Modelling the Long Run. *The Economic Journal, 107*(440), 178-191.

Pesaran, M. H., & Smith, R. P. (1998). Structural Analysis of Cointegrating VARs. *Journal of Economic Surveys, 12*(5), 471-505.

Ravallion, M., & Chen, S. (2007). China's (uneven) progress against poverty. *Journal of Development Economics, 82*(1), 1-42.

Robinson, S. (1976). A Note on the U Hypothesis Relating Income Inequality and Economic Development. *American Economic Review, 66*(3), 437-440.

Saith, A. (1983). Development and distribution: A critique of the cross-country U-hypothesis. *Journal of Development Economics, 13*(3), 367-382.

Sicular, T., Ximing, Y., Gustafsson, B., & Shi, L. (2007). The Urban-Rural Income Gap and Inequality in China. *Review of Income and Wealth, 53*(1), 93-126.

Tao, R., Lin, J. Y., Liu, M., & Zhang, Q. (2004). Rural Taxation and Government Regulation in China. *Agricultural Economics, 31*(2-3), 161-168.

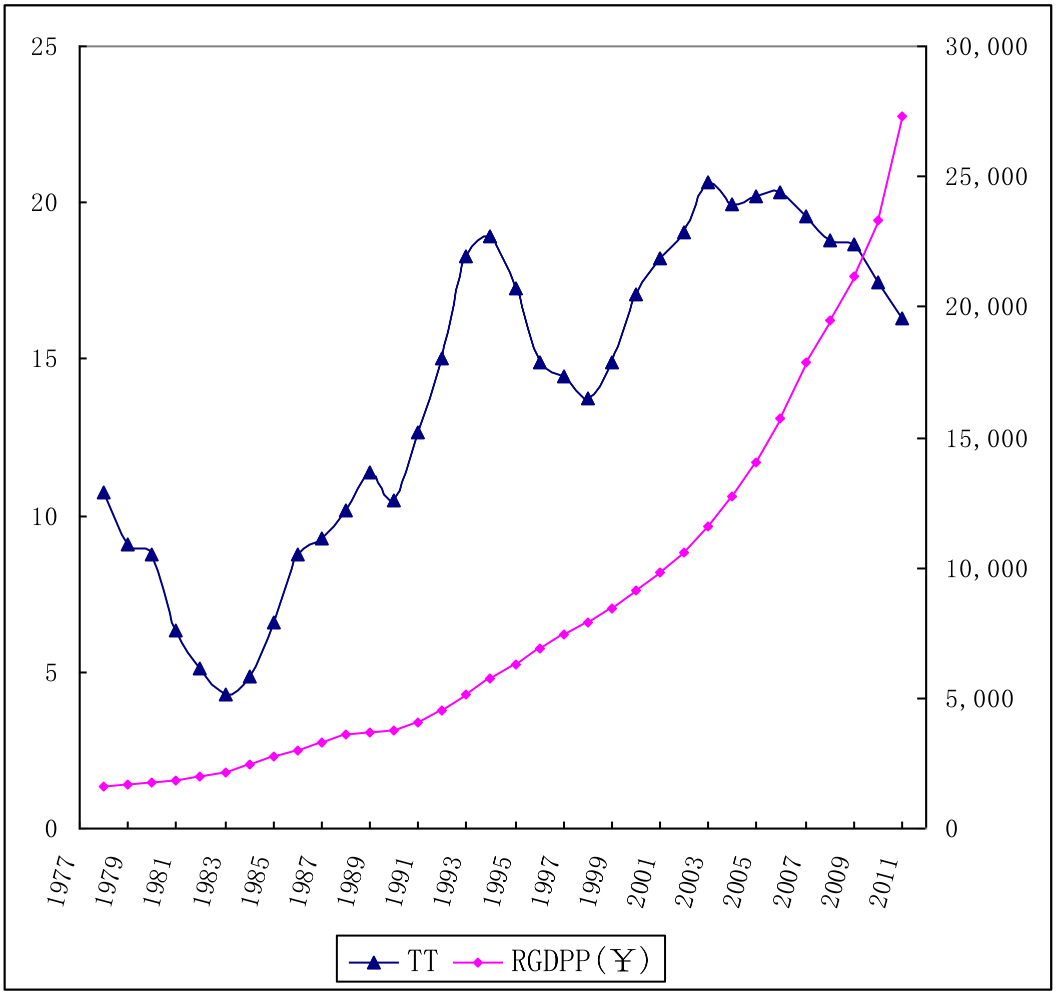
Yang, D. T. (1999). Urban-Biased Policies and Rising Income Inequality in China. *American Economic Review, 89*(2), 306-310.

Yang, D. T. (2002). What Has Caused Regional Inequality in China? *China Economic Review, 13*(4), 331-334.

Zhang, H., Chen, W., & Zhang, J. (2012). Urban–rural Income Disparities and Development in a Panel Dataset of China for the Period from 1978 to 2006. *Applied Economics 44*(21), 2717-2728.

Zhang, Q. and Zou, H.-f. (2012). Regional Inequality in Contemporary China. *Annals of Economics and Finance, 13*(1), 119-143.

Figure 1. Real GDP per capita (at constant 2005 prices) and Inequality 1978-2011



Data source: *China Compendium of Statistics 1949-2009*, and 2010-2012 issues of *China Statistical Yearbook*.

Table 1. Definitions of variables

Variables	Definitions
TT	Theil's index x 100 Calculated by the authors (see appendix 1 for details of the calculation).
URBAN	Urbanization $= \frac{\text{Urban population}}{\text{Total population}} \times 100$
DUAL	Dualism $= \frac{\text{Average productiviy}}{\text{Agriculatural productivity}}$ $= \frac{\text{GDP/Total employment}}{\text{Primary industry GDP/Primary industry employment}}$
INF	Consumer price index series (preceding year = 100)
HEDU	Higher education $= \frac{\text{Enrollment in higher education institutions}}{\text{Total population}} \times 10{,}000$
SEDU	Secondary education $= \frac{\text{Enrollment in secondary schools}}{\text{Total population}} \times 100$
T1992	Dummy variable = 0 for years 1978-1992; 1 for years 1993-2011.

Table 2a. Unit root tests: national time series 1978-2011

Level Variable	ADF test Statistic (Prob[a].)	(C,T,L)[b]	First Difference Variable	ADF test Statistic (Prob[a].)	(C,T,L)[b]
ln(TT)	0.1783 (0.7317)	(0,0,1)	Δ ln(TT)	-3.0492** (0.0034)	(0,0,0)
ln(URBAN)	-3.9926 (0.9999)	(0,0,1)	Δln(URBAN)	-1.7748* (0.0723)	(0,0,0)
ln(DUAL)	-3.1443 (0.1132)	(C,T,0)	Δ ln(DUAL)	-5.0475*** (0.0000)	(0,0,0)
ln(INF)	-0.1289 (0.6312)	(0,0,2)	Δ ln(INF)	-5.6106*** (0.0000)	(0,0,1)
ln(HEDU)	2.1882 (0.9916)	(0,0,1)	Δ ln(HEDU)	-2.3549** (0.0201)	(0,0,0)
ln(SEDU)	-2.7497* (0.0770)	(C,0,1)	Δ ln(SEDU)	-2.5619** (0.0121)	(0,0,0)

a. MacKinnon (1996) one-sided p-values.
b. C,T, L refer to intercept, trend, lag length specified in the ADF tests.
*, **, *** indicate statistical significance at 10%, 5%, and 1% levels respectively.

Table 2b. Unit root tests: provincial panel data 1978-2011

Variable	Test Method[a]	Level Test		First Difference test	
		Statistic	Prob.	Statistic	Prob.
ln(TT)	LLC	-0.0311	0.4876	-16.8175***	0.0000
	ADF-F	80.2041	0.0598	404.841***	0.0000
	PP-F	87.7890	0.0173	506.097***	0.0000
ln(URBAN)	LLC	-0.4840	0.3142	-12.5758***	0.0000
	ADF-F	56.0894	0.6873	514.133***	0.0000
	PP-F	51.1018	0.8369	552.259***	0.0000
ln(DUAL)	LLC	2.5609	0.9948	-26.7957***	0.0000
	ADF-F	17.3154	1.0000	731.494***	0.0000
	PP-F	14.4106	1.0000	798.326***	0.0000
ln(INF)	LLC	-9.59588***	0.0000	-29.2149***	0.0000
	ADF-F	174.278***	0.0000	794.358***	0.0000
	PP-F	794.358***	0.0000	1264.27***	0.0000
ln(HEDU)	LLC	4.3626	1.0000	-15.3012***	0.0000
	ADF-F	19.2056	1.0000	346.461***	0.0000
	PP-F	24.8381	1.0000	371.812***	0.0000
ln(SEDU)	LLC	-1.1128	0.1329	-16.5288***	0.0000
	ADF-F	148.877	0.0000***	352.485***	0.0000
	PP-F	84.8473	0.0286**	294.883***	0.0000

a. LLC, ADF-F, PP-F refer to Levin,Lin & Chu, ADF-Fisher Chi-square, PP-Fisher Chi-square unit root test methods for panel data respectively.
*, **, *** indicate statistical significance at 10%, 5%, and 1% levels respectively.

Table 3. Variable addition tests for the existence of a long relationship[1]

Corresponding Empirical Model	Lagrange Multiplier Statistic [Prob.]	Likelihood Ratio Statistic [Prob.]	F Statistic [Prob.]
Equation (2a)	15.3522** [.018]	21.1933*** [.002]	2.4528* [.074]
Equation (2b)	14.0056** [.030]	18.8688*** [.004]	1.6054 [.235]

Note: This is step 2 of the ARDL estimation procedure in our time series analysis as described section 3. It is a joint test of zero restrictions on the coefficients of additional lagged level variables. The Lagrange Multiplier Statistic, the Likelihood Ratio Statistic and the F Statistic reported indicate that the null hypothesis of no co-integration can be rejected.

*, **, *** indicate statistical significance at 10%, 5%, and 1% levels respectively.

Table 4. Estimation results

DATA TYPE		National time series		Provincial panel data	
Model NO.		Model 1	Model 2	Model 3	Model 4
Estimate method		ARDL-ECM	ARDL-ECM	GMM	GMM
Sample size (adjusted)		31	31	1013	1023
Dependent variable		LNTT	LNTT	LNTT	LNTT
Estimated Long Run Coefficients	Intercept	-48.6562*** (7.2664) [.000]	-69.9779* (33.5823) [.051]	-15.5751*** (4.1392) [.000]	-13.6295*** (2.9726) [.000]
	T1992	-0.2107* (.1131) [.077]	0.0710 (0.1602) [0.663]	0.4570*** (0.1074) [.000]	0.3460*** (0.0754) [.000]
	ln(URBAN)	25.8252*** (3.5423) [.000]	34.1197* (16.9210) [.058]	4.4643*** (0.7840) [.000]	5.3664*** (0.5907) [.000]
	$(\ln(URBAN))^2$	-4.0138*** (.5404) [.000]	-4.7460* (2.3997) [.063]	-0.6512*** (0.1204) [.000]	-0.7430*** (0.0915) [.000]
	ln(DUAL)	2.8331*** (.4304) [.000]	0. 0821 (.7605) [.915]	0.6245*** (0.1746) [.000]	1.0992*** (0.1825) [.000]
	ln(INF)	1.1581** (.5061) [.033]	2.0836* (.8952) [.031]	1.9046* (0.9951) [.056]	1.3227* (0.7108) [.063]
	ln(HEDU)	.6103*** (.1295) [.000]		0.1077 (0.0936) [.250]	
	ln(SEDU)		1.1676 (.9722) [.245]		-0.7684*** (0.2319) [.000]
Error Correction coefficient	ECM_{t-1}	-.4671*** (.0512) [.000]	-.2779*** (.0488) [.000]	—	—
Diagnostic test	LM test for correlation	.2041 [.651]	. 9670 [.325]	—	—
	Ramsey's RESET test	8.8975*** [.003]	12.5780*** [.000]	—	—
	Heteroskedasticity test	1.0811 [.298]	1.0399 [.276]	—	—
	F-stat.	358.38***	298.08***	—	—
	J-statistic	—	—	5.56E-13	6.85E-15
	R-Bar-Squared	.9916	.9909	0.7556	0.7690
ARDL Order selected based on Schwarz Bayesian Criterion		ARDL (3,1,0,0,0,0)	ARDL (2,2,0,1,0,0)	—	—
Adjusted sample range		1981-2011	1981-2011	1979-2011	1979-2011
Adjusted sample size		31	31	1013	1023

Note: Figures in （） are standard errors; figures in [ ] are probability values associated with the obtained T-ratios. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels respectively.

Appendix 1. Calculation of the Theil index

The Theil index has its origin in Shannon's (1948) information theory. Theil (1967) adapted Shannon's formula of expected information content to measure inequality, leading to the now well-known Thei's TT (Conceicao and Galbraith, 2000):

$$TT = \sum \frac{y_i}{Y} \ln \frac{y_i / Y}{1/n} = \frac{1}{n} \sum \frac{y_i}{\mu} \ln \frac{y_i}{\mu} \qquad \text{(A1)}$$

where *n* is the number of individuals in the population, *Y* is the total income of the population, y_i is the income of individual *i*, μ is the average income of the population.

The Theil index can be understood as a summary statistic that measures the extent to which the distribution of income across groups differs from the distribution of population across the same groups (Conceição & Ferreira, 2000). Groups that have higher income shares than their population shares contribute positively to the Theil index; those that have lower income shares than their population shares contribute negatively. If each groups has their "fair" share of income (i.e., each group has the same share of income as its share of population), the Theil index is at its minimum value of zero.

If we consider a population that is divided into *i* groups each with j subgroups, the Theil index can be written as:

$$TT = \sum_i \sum_j \frac{Y_{ij}}{Y} \ln(\frac{Y_{ij}}{Y} \Big/ \frac{N_{ij}}{N}) \qquad \text{(A2)}$$

where Y_{ij} is the income of subgroup *j* in group *i*; N_{ij} is the population size of subgroup *j* in group *i*.

To calculate the national time series Theil index given provincial data of China, we rewrite equation (A2) as:

$$TT = \sum_i \sum_j \frac{Y_{ij}}{Y} \ln \left[\frac{\frac{Y_{ij}}{Y}}{\frac{N_{ij}}{N}} \right] = \sum_i \sum_j \frac{N_{ij}}{N} \frac{\overline{Y}_{ij}}{\overline{Y}} \ln \left[\frac{\frac{Y_{ij}}{Y}}{\frac{N_{ij}}{N}} \right] = \sum_i \sum_j \frac{N_{ij}}{N} \frac{\overline{Y}_{ij}}{\overline{Y}} \ln(\frac{\overline{Y}_{ij}}{\overline{Y}}) \qquad \text{(A3)}$$

where *i*=1, 2 representing the urban area and rural area, respectively; *j*=1, 2,⋯， 31, representing 31 provinces (including autonomous regions and directly-administered municipalities); N_{ij} is the urban (i=1) or rural (i=2) population in province j; N is the total population of China; $\overline{Y}_{ij}$ is the average urban or rural income in province j; $\overline{Y}$ is the average income in China.

To calculate the provincial panel Theil index, we rewrite equation (A2) to

$$TT = \sum_{i=1}^{2} \frac{Y_i}{Y} \ln \frac{Y_i / Y}{N_i / N}$$

where Y_1 = total annual disposable income of urban households

Y_2 = total annual net income of urban households

$Y = Y_1 + Y_2$

N_1 = urban population

N_2 = rural population

$N = N_1 + N_2$

References

Conceição, P., & Ferreira, P. (2000). The Young Person's Guide to the Theil Index: Suggesting Intuitive Interpretations and Exploring Analytical Applications. *UTIP Working Paper Number 14*.

Conceicao, P., and Galbraith, J. K. (2000). Constructing Long and Dense Time-Series of Inequality Using the Theil Index. *Eastern Economic Journal, 26*(1), 61-74.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*, 379-423.

Theil, H. (1967). *Economics and Information Theory*. Chicago: Tand McNally and Company.

Appendix 2. National Time Series Data

Year	TT	RGDPP	URBAN	DUAL	INF	HEDU	SEDU	T1992
1978	10.7674	1582	17.9152	2.5020	100.70	3.52	3.86	0
1979	9.0501	1679	18.9611	2.2324	101.90	4.73	4.02	0
1980	8.7523	1789	19.3911	2.2783	107.50	5.42	4.83	0
1981	6.3032	1859	20.1565	2.1361	102.50	6.03	6.23	0
1982	5.1147	1999	21.1305	2.0405	102.00	6.58	7.14	0
1983	4.2832	2185	21.6236	2.0218	102.00	8.89	6.84	0
1984	4.8644	2481	23.0143	1.9932	102.70	10.46	6.10	0
1985	6.5710	2779	23.7069	2.1945	109.30	11.59	5.63	0
1986	8.7494	2979	24.5249	2.2456	106.50	12.78	4.90	0
1987	9.2491	3270	25.3193	2.2374	107.30	11.35	4.52	0
1988	10.1497	3584	25.8147	2.3098	118.80	11.72	4.39	0
1989	11.3821	3671	26.2102	2.3920	118.00	13.38	4.53	0
1990	10.5001	3755	26.4097	2.2164	103.10	16.09	4.66	0
1991	12.6606	4046	26.9402	2.4341	103.40	17.49	4.79	0
1992	15.0195	4565	27.4599	2.6847	106.40	17.92	4.77	0
1993	18.2827	5143	27.9901	2.8617	114.70	18.61	4.54	1
1994	18.9025	5751	28.5098	2.7340	124.10	18.47	4.29	1
1995	17.2399	6311	29.0404	2.6149	117.10	18.04	4.27	1
1996	14.8686	6870	30.4799	2.5646	108.30	17.65	4.32	1
1997	14.4210	7432	31.9100	2.7287	102.80	18.64	4.36	1
1998	13.7502	7937	33.3502	2.8366	99.20	21.4	4.30	1
1999	14.8848	8469	34.7797	3.0418	98.60	23.35	4.50	1
2000	17.0732	9111	36.2198	3.3194	100.40	23.99	4.80	1
2001	18.2303	9796	37.6597	3.4743	100.70	24.68	5.08	1
2002	19.0716	10614	39.0898	3.6383	99.20	25.67	5.28	1
2003	20.6257	11605	40.5302	3.8367	101.20	27.32	5.48	1
2004	19.9720	12700	41.7600	3.5018	103.90	32.87	5.81	1
2005	20.2312	14053	42.9900	3.6611	101.80	43.88	5.88	1
2006	20.3502	15747	43.9002	3.7554	101.50	56.34	6.21	1
2007	19.5674	17883	44.9402	3.6672	104.80	70.33	6.52	1
2008	18.8002	19506	45.6823	3.5019	105.90	85.79	6.70	1
2009	18.6684	21198	48.3417	3.6871	99.30	102.59	6.73	1
2010	17.4398	23299	49.9497	3.6354	103.30	119.44	7.82	1
2011	16.3288	27309	51.2703	3.4655	105.40	132.28	7.82	1

Data Source: *China Compendium of Statistics 1949-2009*, and 2010-2012 issues of *China Statistical Yearbook*.