Research Paper No. 2007/05

Regional Income Inequality in Rural China, 1985-2002

Trends, Causes and Policy Implications

Guanghua Wan*

January 2007

Abstract

This paper depicts the trend of regional inequality in rural China for the period 1985-2002. The total inequality is decomposed into the so-called within- and between-components when China is divided into three regional belts (east, central and west). A regression-based accounting framework is then used to explore root sources of the rising inequality. Policy implications are discussed.

Keywords: China, inequality, spatial decomposition, rural
JEL classification: O18, P2, D63

*UNU-WIDER, Helsinki.

This study has been prepared within the UNU-WIDER project on Inequality and Poverty in China.

UNU-WIDER acknowledges with thanks the financial contributions to its research programme by the governments of Denmark (Royal Ministry of Foreign Affairs), Finland (Ministry for Foreign Affairs), Norway (Royal Ministry of Foreign Affairs), Sweden (Swedish International Development Cooperation Agency—Sida) and the United Kingdom (Department for International Development).

ISSN 1810-2611 ISBN 92-9190-944-0 ISBN 13 978-92-9190-944-5

UNU World Institute for Development Economics Research (UNU-WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Typescript prepared by Lorraine Telfer-Taivainen at UNU-WIDER

1 Introduction

It is widely recognized that regional inequality in China has been on the rise since economic reforms were initiated in the late 1970s (Kanbur and Zhang 2005; Wan 2005). In addition to its repercussions on social and political stability, such a rise has hampered poverty alleviation (Ravallion and Chen 2004; Zhang and Wan 2006) and is found to be detrimental to long-run economic growth (Wan et al. 2006). Many Chinese scholars also consider high inequality as a major contributor to the sluggish domestic demand in China. It is thus not surprising to witness a broad and growing interest in China's regional inequality. Earlier studies largely focused on the measurement of regional inequality. Subsequent efforts were devoted to breakdown total inequality into various components, either by population subgroups (Tsui 1991) or by factor components (Wan 2001). Recently, the technique of regression based decomposition has gained popularity (Fields and Yoo 2000; Morduch and Sicular 2002; Wan 2002) and has been applied to China (Morduch and Sicular 2002; Wan 2004; Wan and Zhou 2005).

Despite a large volume of literature on regional inequality in China, few existing studies constructed time profiles of inequality among rural regions in China.[1] This is surprising given that a dominant proportion of China's population live in the countryside and, as discussed later, rural inequality is a large component of the overall regional inequality. More importantly, rural inequality is fundamentally different from the urban counterpart in terms of causes, trends and policy implications. For example, geography is much more important in driving rural inequality than the urban inequality. When encompassing weather, infrastructure and other natural resources, the geography variable would account for a very significant share of total rural inequality. This is not necessarily the case for the urban sector despite the probable relevance of location as a determinant of wages. Needless to say, distribution of arable land is relevant to rural inequality but not to the urban counterpart in China. Clearly, while studies on whole China or urban China are important, there exist obvious justifications for a separate paper focusing on rural China.

This paper will fill such a gap in the literature by providing a time profile of regional income inequality in rural China for the period 1985-2002. Such a time profile was appealed for by Rozelle (1994). Another purpose of the paper is to identify the components of rural regional inequality. This is accomplished by undertaking conventional as well as regression-based inequality decompositions. These decompositions offer different insights into the determinants of the total inequality. Policy implications, also, will be explored.

---

[1] A search in Econlit using keywords 'China', 'region', 'rural' and 'inequality' produced 59 journal article entries and only a few of them touched on, but did not focused on, rural regional inequality.

The paper is organized as follows. The next section describes the data and the time profile of rural regional inequality, wherein conventional and newly proposed methods will be employed to decompose total inequality into two broad components: between regional belts (i.e., east-central-west China) and within these belts. Section 3 applies the regression-based decomposition to rural China, which helps reveal the root sources of regional inequality. Finally, we discuss policy implications in Section 4.

2 Data and preliminary analysis

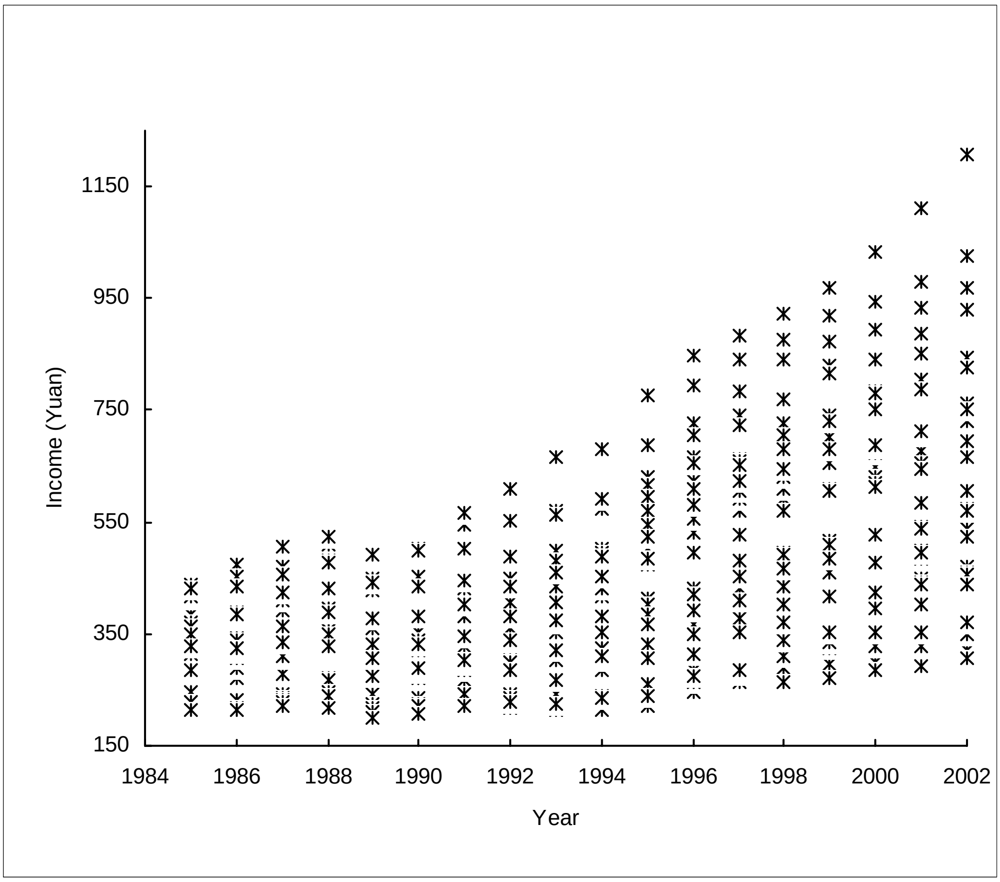
As a precursory note, it is useful to mention that a substantial proportion, in the order of 25 per cent or so, of China's regional inequality is attributable to the urban–rural gap.[2] The remaining is due to inequalities within urban and rural regions. In accounting for the total regional inequality for China as a whole, these so-called within-components are given by their respective Theil-L index estimates, weighted by their population shares. Since the unweighted rural regional inequality is found to be larger than the urban counterpart (Wan 2005) and a dominant percentage of population lives in the rural areas, the contribution of rural regional inequality to the total must be substantial.

To accomplish research objectives of this paper, most of our data are compiled from the *China Rural Household Survey Yearbook* (NBS various years) for the period 1985-2002. Earlier data are incomplete. Ideally, rural population should be used as our income observations are for rural residents. However, we failed to find consistent rural population series for all regions. Instead, agricultural population statistics are used. They are expected to be highly correlated with the rural population and are available from the *China Rural Statistical Yearbook* (NBS various years). Excluding Hong Kong, Taiwan and Macao, there are 31 regions (provinces, autonomous regions or metropolitan cities) in China. However, our sample contains data for 28 regions with Hainan merged with Guangdong, Congqing merged with Sichuan, and Xizang (Tibet) excluded. Data for Tibet are not complete. As argued by Wan (2001), such exclusion is not expected to distort the analytical results. All data in value terms are deflated by regional rural consumer price indices (CPIs) as well as the regional price indices compiled by Brandt and Carsten (2004).

The deflated regional income data are plotted in Figure 1 against years; for each year the plot contains per capita real incomes for all 28 regions. The Figure shows that while real income has been increasing over time, its dispersion is also on the rise as indicated by the expanding height of the plots over time. According to Figure 1, regional incomes had been increasing over 1985-89. After a setback for most regions in 1989-90, the

[2] If differences in price levels and inflations between urban and rural areas were not considered, this proportion would be over-estimated as in Kanbur and Zhang (1999). See Sicular et al. (2007).

Figure 1: Regional per capita real income, by year



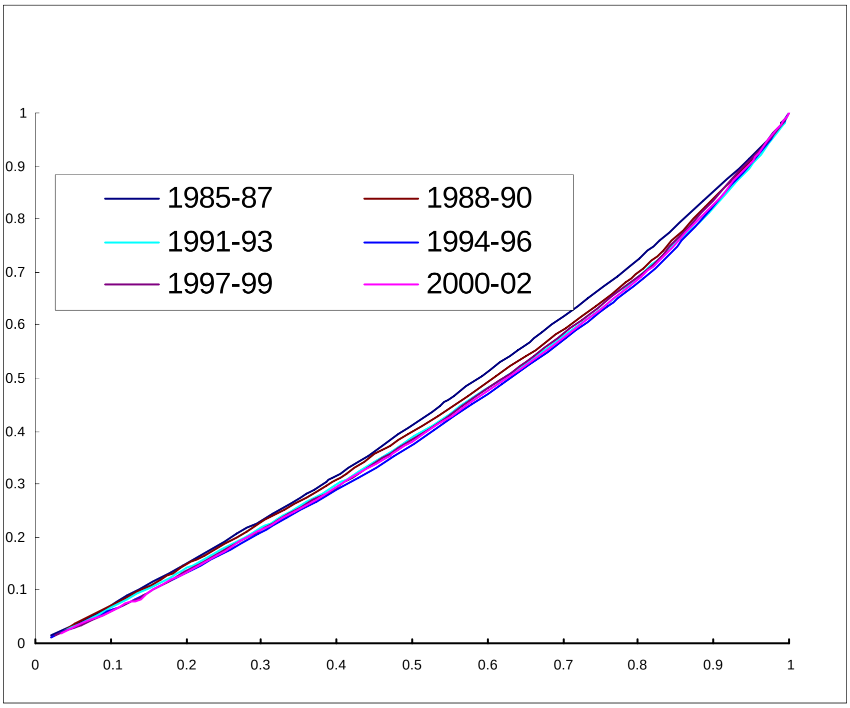
increasing trend resumed. The setback is probably caused by the austerity programme initiated by the Chinese government in 1989 (Wan 2001).

Reasonably assuming no changes in the composition of regions in the rich and poor groups, the poor (lower segments of the plots) consistently experienced slower income growth than the rich (upper segments of the plots). In fact, real income declined or was stagnant before 1996 for the poor regions as a group. Even after a small rise in 1996, this group's income rose little over 1996-2002. In contrast, income increased from 1985 to 1988 for the rich regions. For this group, there was a small drop in income in 1989, but the increasing trend resumed right after and continued strongly until 2002. Judging from these observations, one may conclude that regional income inequality in rural China has been increasing in both absolute as well as relative terms, respectively indicated by the expanding height of the plots over time, and by the differing growth rates for the poor and rich regions.

Examination of Figure 1 reveals that the gaps between the income groups (top, middle, and bottom segments of the plots) seem to have expanded more than those within these groups or segments. This clustering of regional income in recent years implies some forms of polarization in China. In other words, income has been diverging more between income groups than within income groups. Nevertheless, Figure 1 may be misleading as far as gathering inequality trend is concerned because the expansion in income dispersion had been accompanied by changes in income levels. It is known that an identical income growth for all regions can also result in increased dispersions, as in Figure 1, but such growth leaves inequality unchanged as long as relative rather than absolute inequality measures are used.

A formal way to analyze inequality is to construct Lorenz curves and conduct stochastic dominance analysis. For this purpose, Lorenz curves are obtained for each of the 18 years. Although there is a tendency for the Lorenz curve to move downwards over time, any first-degree stochastic dominance is not clearly visible when they are all displayed in one diagram. On the one hand, this may be caused by 'too much information'—many curves are squeezed onto one diagram. On the other, this is understandable as inequality changes are usually small from one year to the next. To reduce distractions caused by 'too much information', we average Lorenz curves over a 3-year interval and present these curves in Figure 2. Unfortunately, even Figure 2 does not exhibit any first-degree stochastic dominance very clearly. As a consequence, pair-wise comparisons of these curves have to be done and they indicate that nine out of the 15 pairs of the Lorenz curves cross, mostly at the top or bottom ends of the distributions.

Figure 2: Lorenz curves



When Lorenz curves cross, they cannot be used to rank income distributions. In this case, second or higher degree of stochastic dominance can be introduced. Alternatively, summary inequality measures could be used instead (Fields 2001). To minimize possible sensitivities to inequality measures,[3] we compute most relative inequality indices that are commonly in use, namely the Gini, Theil-L, Theil-T and half CV^2.[4]

Let Z denote the target variable, μ denote the mean of Z, j index observations (j = 1, 2, …, N), the following formulae can be used: Atkinson = $1-\prod_j \left(\frac{Z_j}{\mu}\right)^{1/N}$, Theil-L = $\frac{1}{N}\sum_j Ln\frac{\mu}{Z_j}$, and Theil-T = $\frac{1}{N}\sum_j \frac{Z_j}{\mu} Ln\frac{Z_j}{\mu}$. The Atkinson index is not considered here because it can be expressed as a monotonic transformation of Theil-L (Shorrocks and Slotjje 2002).

The computed values are tabulated in Table 1 (left panel). Since CV^2 violates the principle of transfer, values in the last column are reported only for comparison purpose as there are many studies in China using the measure CV^2. Results in Table 1 show that all measures are consistent in demonstrating a rising trend in regional inequality in rural China. In particular, the inequality increased rather dramatically until 1995-96. After that, the increasing trend became moderate. This finding is consistent with Figure 1, which shows that income of poor regions experienced little growth before 1995-96 but some improvement afterwards while the rich regions exhibited growth throughout the period. The slowing down in inequality increases after 1995-96 may be caused by the implementation of the grain price support policy which benefited poor regions more (Zhang 2005). Since the trend was only moderated, not really reversed, other forces must be stronger than the policy change in pushing up the long-run inequality trend. Identifying these other forces is crucial for policymakers if rural regional inequality is needs to be brought down.

As a by-product, we calculated inequality values using undeflated data (see the right panel in Table 1). As is expected, not taking into consideration inflation and regional price levels leads to upward biases in inequality measurement. The biases are quite substantial. What is interesting, and perhaps surprising, is that the biases are larger in

---

3 Different measures imply different social welfare functions and different aversions to inequality (see Dagnum 1990).

4 There exist alternative ways to calculate the Gini coefficient. We follow Silber (1989) by defining Gini = P´QI, where P is the vector containing population shares and I is the vector containing income shares, both sorted in ascending order by the per capita income variable. Q is a square matrix with 0 on the diagonal, 1 above the diagonal and -1 below the diagonal.

the early years, a finding consistent with Brandt and Carsten (2004). Also, the biases are less severe when using Gini relative to other measures, possibly due to the differing sensitivities of these measures to different sections of the underlying Lorenz curves.

Table 1: Regional inequality in rural China

	Deflated data				Undeflated data			
	Gini	Theil-L	Theil-T	CV^2	Gini	Theil-L	Theil-T	CV^2
1985	0.109	0.020	0.019	0.037	0.152	0.038	0.042	0.095
1986	0.123	0.025	0.024	0.047	0.171	0.047	0.050	0.114
1987	0.129	0.027	0.026	0.052	0.180	0.052	0.056	0.127
1988	0.134	0.029	0.028	0.057	0.187	0.056	0.061	0.138
1989	0.137	0.030	0.029	0.059	0.194	0.060	0.065	0.148
1990	0.141	0.032	0.031	0.062	0.198	0.063	0.069	0.162
1991	0.142	0.032	0.032	0.065	0.208	0.070	0.078	0.185
1992	0.151	0.036	0.036	0.072	0.215	0.074	0.082	0.194
1993	0.164	0.043	0.042	0.086	0.231	0.084	0.094	0.226
1994	0.170	0.046	0.045	0.090	0.228	0.082	0.089	0.210
1995	0.186	0.056	0.054	0.107	0.234	0.085	0.092	0.211
1996	0.188	0.058	0.055	0.109	0.221	0.076	0.080	0.180
1997	0.186	0.057	0.054	0.106	0.214	0.072	0.076	0.169
1998	0.186	0.056	0.053	0.105	0.208	0.067	0.071	0.158
1999	0.188	0.058	0.055	0.109	0.212	0.070	0.074	0.164
2000	0.188	0.057	0.055	0.110	0.219	0.075	0.080	0.181
2001	0.195	0.061	0.059	0.118	0.225	0.079	0.084	0.190
2002	0.195	0.062	0.059	0.119	0.226	0.080	0.085	0.194

3 Spatial decomposition

To assess if the rising inequality is due to enlarged gaps between regional belts or within regional belts, we undertake subgroup decomposition following Shorrocks (1984). As in most studies, we divide China into three belts. The central belt includes Shanxi, Guangxi, Inner Mongolia, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei and Hunan. The west belt includes Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang. The east belt consists of the remaining regions, all along the coast. The inequality index we use is the Theil-L coefficient. Other inequality measures are not appropriate for the conventional subgroup decomposition; see Shorrocks and Wan (2005).

The decomposition results are shown in Figure 3 (numerical values are tabulated in Appendix Table A1). It is found that (a) all individual components (within each belt and between belts) exhibit increasing trends, clearly demonstrating income divergence within belts as well as divergence between belts; (b) the divergence between belts expanded faster than divergence across regions within belts, resulting in a growing share of the between-component; (c) there is an oscillatory cycle around the total

Figure 3A: Traditional east-central-west decomposition: Theil-L

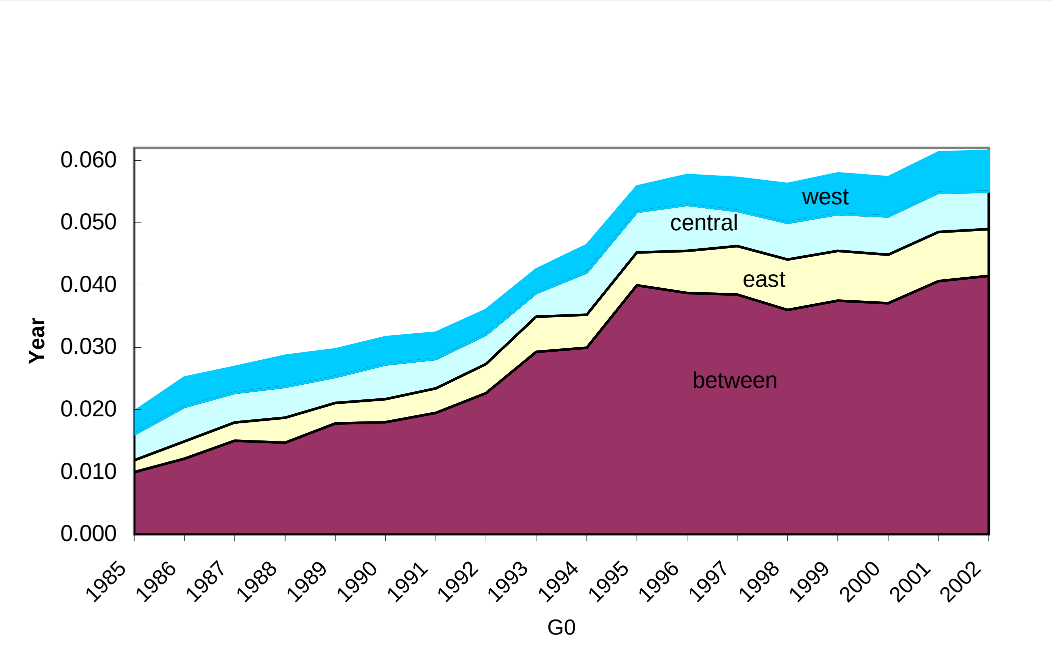


Figure 3B: East-central-west decomposition: Theil-T

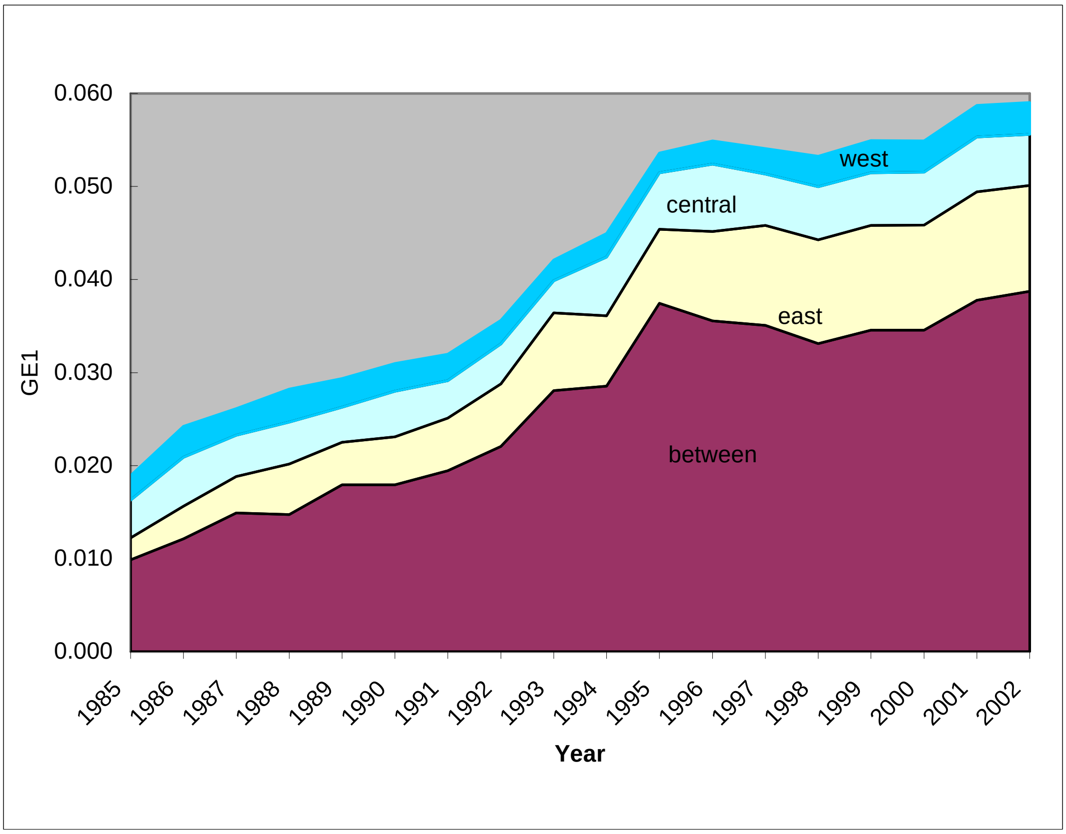


Figure 3C: East-central-west decomposition: Gini

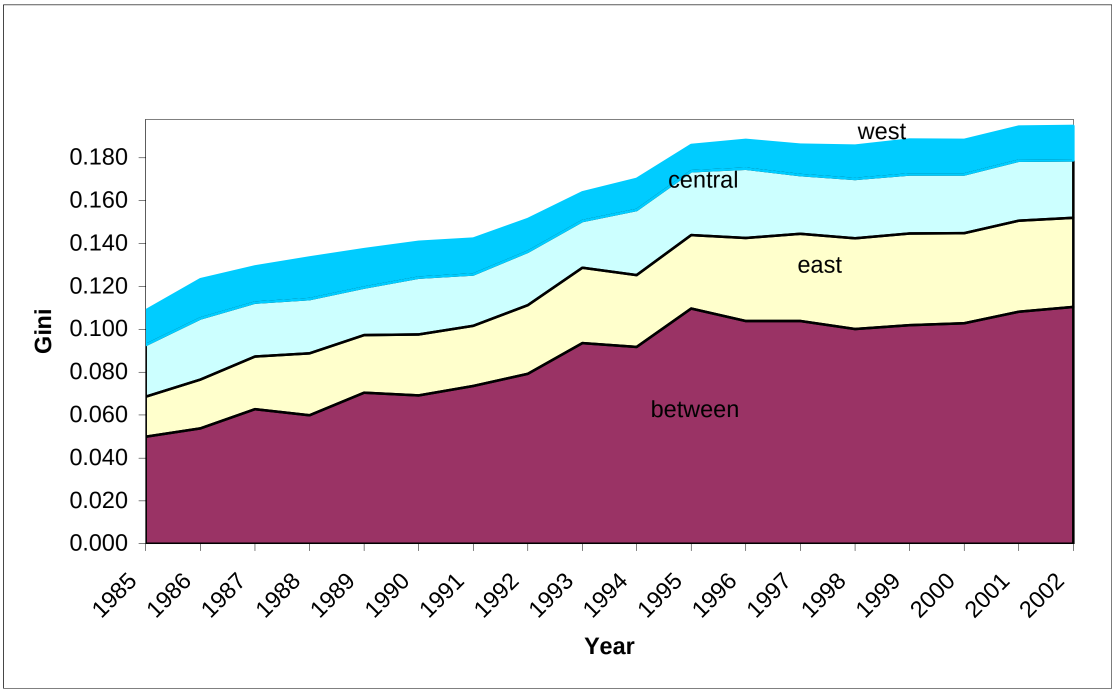


Figure 3D: East-central-west decomposition: CV2

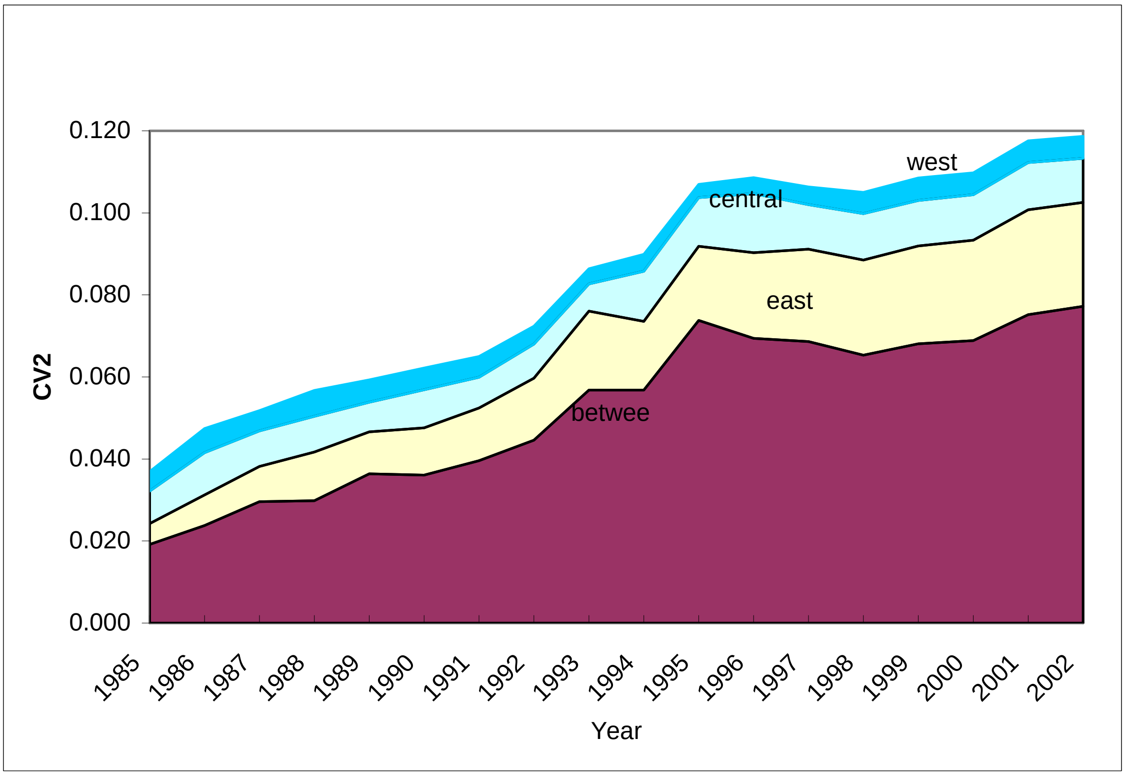
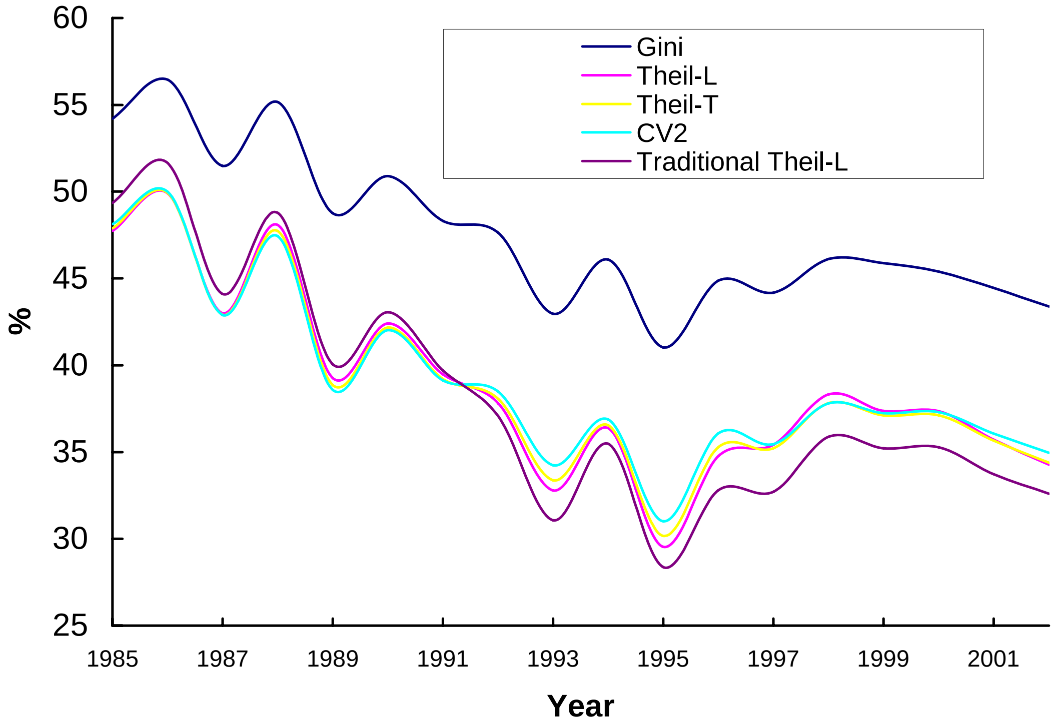


Figure 4: Overall within-belt contribution to regional inequality (%)



inequality trend and it is largely driven by the between-component. This implies that there are forces driving inter-belt income divergence, and they fluctuate from year to year. These forces may include cyclical weather conditions, the 'cobweb phenomena' often present in the agricultural sector, and biological cycles of perennial crops or plants. For example, differences in production structure across the belts may contribute to the cyclical pattern. The west and central produce more grain and fruits. Fruits typically have a bumper year followed by a lean year and this contributes to the income correlation between regions within these belts. In a bumper year, the income gaps between the coast and inland regions will narrow, otherwise they will expand. Weather and market conditions also vary from year to year. In bad years, regions producing similar products will suffer income drops together, leading to larger income gaps between belts. In good years, these gaps may narrow down; (d) the year 1995 deserves special attention. In this year, total inequality jumped more than usual and all components seem to have reached their peak values. After 1995, the between-component stayed more or less constant but all within-components continued to rise, particularly that within the west belt. It would be interesting to explore if this is related to the major taxation reform implemented in 1994. Also, grain price support policy was introduced in 1995. Such a nationwide policy shift certainly helped raise income levels of the poor regions more, leading to smaller gaps between regional belts.

One can add up all the within-components and compute their overall percentage contribution to total inequality. Figure 4 confirms the early observation that the total within-belts contribution has declined in relative terms although the total absolute contribution has not. It is also clear that the declining trend was contained from 1995 to 1998 and resumed after 1999. Given the finding that the between-component is a positive function of the number of groups involved in subgroup decompositions (Shorrocks and Wan 2005), such a large and increasing between-component is rather surprising. Here, there are only three groups and yet the between-component is so large, which must imply very substantial inter-belt income gaps. Note that current literature points to a small between-component with only a few exceptions (Shorrocks and Wan 2005).

Looking into the individual within-components (see Appendix Table A1), dispersions within the west belt were rather small and those within the central belt were large until the late 1990s and early 2000s. The within-component of the east belt was moderate in the 1980s. It increased to a level more or less compatible to that of the central belt around the mid 1990s. Since then, the contribution of the east belt dominated the total within-component. Thus, as far as the within components are concerned, the east is more important. It is beyond the scope of this paper to explore why the east regions became least homogeneous while the west regions were always more homogeneous.

What about sensitivity of the above decomposition results to inequality measures? Previously, it has been uncommon or inappropriate to use inequality measures other than the Theil-L index for subgroup decomposition. Therefore, examining sensitivity in this context has so far been rare or not possible. In what follows, we propose a subgroup decomposition procedure which can be used with any inequality measures. We then apply this approach to the Chinese data. Empirical outcomes will be compared with the earlier results.

Our approach is inspired by the Shapley procedure of Shorrocks (1999). Briefly put, applying this procedure requires a function between the target variable such as income and its determinants. Expressing the function as Y = f(X), one can then apply any operator to both sides of the equation and attribute the total value, as defined by the operator, to the contributions made by individual elements in X. For example, given an income generating function, one can apply inequality operator to both sides of the function, and attribute total income inequality to various components associated with income determinants. Since an identity is a special function, we will rely on a defined identity to decompose regional inequality into the between-belt and within-belt components.

The identity we define must express regional income as a function, being linear or non-linear, of income gaps between regional belts and income gaps across regions within belts. This can be achieved by defining u_i as the average income of, and d_i as the dummy variable for, belt *i*. Also, let δ_1 denote deviations of per capita income of eastern regions from the average income of the east belt. Similarly we can define δ_2 and δ_3. Now, regional income Y can be written as

$$Y = u_i + d_1\delta_1 + d_2\delta_2 + d_3\delta_3 \quad (1)$$

It is noted that u_i take identical values for those regions belonging to the same belt. Therefore, it can be used to represent income gaps between belts. The other three terms in (1) capture income gaps across regions within individual belts.

We can proceed by constructing various counterfactuals. Assume absence of income gaps between belts (i.e., all three belts have the same mean income, say national average, as denoted by u) as well as absence of income gaps within belts. This is equivalent to replacing u_i by u and $d_1\delta_1$, $d_2\delta_2$, $d_3\delta_3$ by 0. Substituting these into (1) will produce identical income for every region. In this case, regional inequality is zero. Now, permitting presence of income gaps between belts only (i.e., u_i take their defined values) while keeping income gaps within individual belts absent (i.e., $d_1\delta_1 = d_2\delta_2 = d_3\delta_3 = 0$). We can substitute these values into (1) and calculate the corresponding inequality. This inequality is only caused by income gaps between belts, not within belts. By the same token, we can replace u_i by u and any two of $d_1\delta_1$, $d_2\delta_2$, $d_3\delta_3$ by 0 (only one of them taking its defined values), the corresponding inequality must be due to income gaps within the relevant belt. This kind of counterfactual can be constructed under all possible combinations of replacements. Alternative estimates of the same inequality component can be obtained; they are averaged to reach the final estimate. This is essentially what Shapley decomposition does (see Shorrocks 1999; Wan and Zhou 2005).

Applying the above decomposition to Chinese data produces results that are quite consistent with the earlier decomposition results (see Table A2 in the Appendix). For comparison purpose, we plot the between-component in Figure 4 as well. The plot resembles the traditional decomposition results quite well, indicating the robustness of earlier decomposition results. For example, they show a clear declining trend in the percentage contribution of the overall within-components. As the number of regions in each belt is kept constant for the decomposition, the faster increases in the between-component are indicative of enlarging income gaps between belts more than those within belts, confirming the early finding of polarization. Interestingly, the results are slightly different when the Gini index is used. The other three give almost indistinguishable percentage contributions. According to the Gini decomposition, the within-component was almost 50 per cent in 1985, and became smaller over time reaching 45 per cent in 2002. The other indicators show similar declining trends, but starting with 45 per cent in 1985 declining to 35 per cent in 2002.

4 Sources of rural regional inequality

Given the consistent rises in regional inequality in rural China, one naturally wonders what factors drive this trend. It is not difficult to speculate about possible determinants of regional inequality in China. What is more interesting and challenging is to quantify the contributions of these determinants. Kanbur (2002) appealed for linking inequality with fundamental variables because simply breaking down total inequality into the usual within- and between-components (as done in the preceding section) is insufficient. To quantify contributions of various determinants to total inequality, we follow the regression-based decomposition approach of Wan (2004), which has a number of advantages. In particular, it does not depend on inequality measures; it is applicable irrespective of functional form for the regression equation, and it permits interactive terms of independent variables. Interested readers are referred to Wan (2002).

Intuitively, if every region possesses the same amount of every input as well as same returns to factors, there would be no inequality. Same returns may not be realistic, but has been presumed in most, if not all, previous studies. We will deal with differing returns in a separate paper. Assuming same returns, it is the spatial distributions of factor inputs that matter. Postulating, without loss of generality, that the marginal impact of a factor is positive on income generation, its dispersion would contribute positively to inequality if the input variable is positively correlated with total income. On the contrary, if rich regions possess less of this factor than poor regions, it would help reduce regional inequality. In the unlikely case of a linear income function $Y = \Sigma \beta_i X_i + \beta_0$, it is possible to express the Gini index of income as a weighted sum of concentration indices of factor inputs:

$$\text{Gini}(Y) = \Sigma \beta_i E(X_i)/E(Y)\, C(X_i) \quad (2)$$

where β_i denotes marginal income of X_i, E is the expectation operator, and $C(X_i)$ denotes the concentration index of X_i. Resembling the Gini coefficient, $C(X_i)$ is a measure of dispersion of X_i. It takes values between 1 and -1. Broadly speaking, if X_i and Y are positively correlated, $C(X_i) > 0$. Otherwise, $C(X_i) < 0$. Although equation (2) may not be derivable under other inequality indices or when the income function is not strictly linear, it does help demonstrate our point that total inequality can be accounted for by dispersions of factor inputs.

Turning to empirical factor-income relationships in China, we plot income against each of the input variables—see Figure 5. The trend lines in the figure indicate that capital, schooling, and industrialization are positively correlated with income while land is perhaps unrelated to income. Figure 5D depicts a non-linear curve between income and schooling or education. To a lesser extent non-linearity also appears in Figure 5B where the correlation does not seem to exist. Of course, such bivariate correlations may not reveal the true relationship as other factors are not controlled for. Also, the impact of factor dispersions on total inequality cannot be directly discerned from Figure 5. To account for rural regional inequality in China, we now turn to the regression-based decomposition technique.

Figure 5A: Income and capital

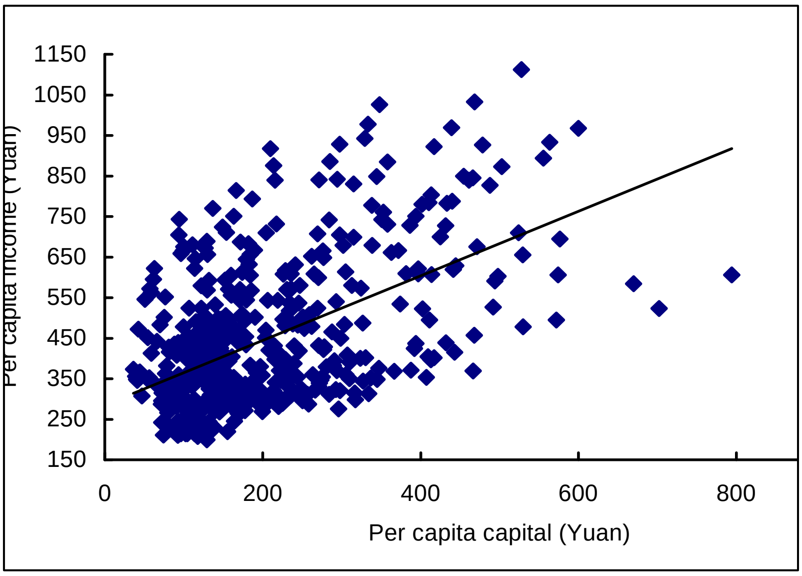


Figure 5B: Income and land

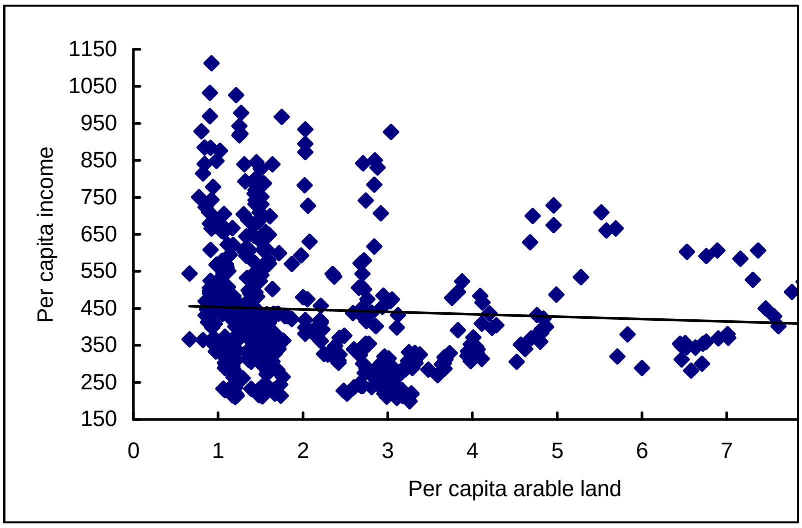


Figure 5C: Industrialization and income

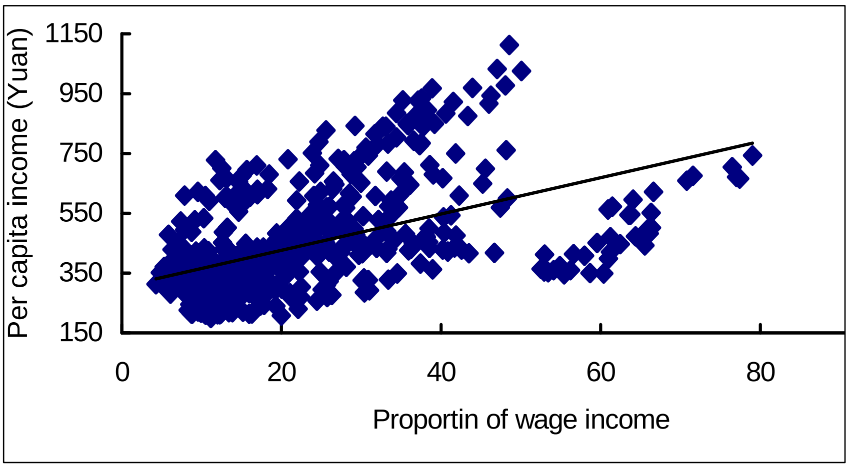
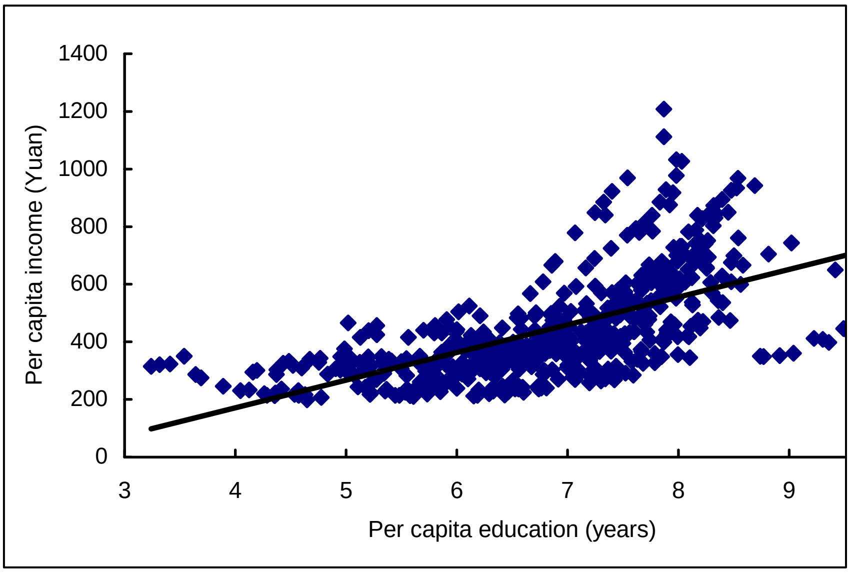


Figure 5D: Education and income



Essentially, the regression based decomposition combines a regression model with the Shapley value framework of Shorrocks (1999). The basic idea is to attribute inequality in the dependent variable to contributions of the residual and independent variables in the regression equation. Following the before-after principle of Cancian and Reed (1998), Wan (2004) proposed to define the contribution of the residual by I(Y) – I($\hat{Y}$), where $\hat{Y}$ is the predicted value based on the estimated regression model and I denotes any inequality measure. To obtain the contributions of independent variables to I($\hat{Y}$), the Shapley procedure developed by Shorrocks (1999) can be applied. The procedure is fully founded on the co-operative game theory and is applicable irrespective of the functional form or inequality measure—for technical details, see Shorrocks (1999).

To estimate the empirical income-generation function, we start with the human capital theory which dictates that income is a function of education and experience. At the aggregate regional level, it is not possible to define experience or its proxy. Further, rural income also depends on production inputs such as land, capital and so on. Consequently, we consider the following variables:

Y = per capita net income (*yuan*/head);
K = capital stock (1000 *yuan*/head);
Land = arable land (*mu*/head, 1 *mu* = 1/15 Ha);
DEP = dependency ratio;
HH = household size;
IND = degree of industrialization = wage income/net income;
EDU = average years of schooling of working age members (years/head)

Note that DEP and HH are included to control for labour input since these three variables are linearly related, thus inclusion of any two of the three is sufficient. In addition, regional dummy variables (to account for geographic location and location-related factors) and dummy variables for years 1992 (marking Deng Xiaoping's tour of southern China), 1995 (marking the start of the grain price support policy) are included. Also, a time trend variable is incorporated to control for possible changes in macro-economic environment and technology or other shifts over time.

Regarding functional form, the conventional practice is to specify a log-linear form, the so-called Mincer function. We experimented with log-linear, linear-linear, double-log and linear-log (i.e., the dependent variable is untransformed but independent variables are in logarithms). Given our panel data model, the disturbance term is proposed to be heteroscedastic across regions and autoregressive over time for individual regions. Denote the disturbance term by ε_{it}, the error process satisfies:

$$\begin{aligned}
&\varepsilon_{it}=\rho_i\varepsilon_{it-1}+u_{it} \text{ for all } i \text{ and } t,\\
&E(u_{it})=0 \text{ for all } i \text{ and } t,\\
&E(u_{it}^2)=\sigma_i^2(1-\rho_i^2) \text{ for all } i,\\
&E(u_{it}u_{js})=0 \text{ for } i\neq j \text{ or } t\neq s,
\end{aligned}$$

Consequently, the variance-covariance matrix becomes (see Kmenta 1986)

$$\begin{bmatrix} R_1\sigma_1^2 & 0 & \dots & 0 \\ 0 & R_2\sigma_2^2 & . & 0 \\ . & . & . & . \\ 0 & 0 & \dots & R_N\sigma_N^2 \end{bmatrix}, \text{ where } R_i = \begin{bmatrix} 1 & \rho_i & \rho_i^2 & . & \rho_i^{T-1} \\ \rho_i & 1 & \rho_i & . & \rho_i^{T-2} \\ . & . & . & . & . \\ . & . & . & . & . \\ \rho_i^{T-1} & \rho_i^{T-2} & . & \rho_i & 1 \end{bmatrix},$$

ρ_i denotes the correlation coefficient between successive errors for region i. The estimation can be easily implemented using the econometric software Shazam.

Estimation results for the income-generation functions are presented in Table 2 (dummy variable and time trend terms are not reported but are available from the author upon request). The usual χ^2 test can be employed to choose between the log-linear and double-log models, and between the linear-linear and linear-log models. The test statistic is given by twice the absolute difference in the log-likelihood values of relevant models, with degrees of freedom equal to the difference in the number of parameters in theses models. Since all the models have the same number of parameters, one can simply compare the relevant log-likelihood values and make a choice. As a consequence, the linear-linear and log-linear models are preferred.

Table 2: Estimation results of income functions

	Estimate	T-ratio	p-value	Estimate	T-ratio	p-value
	log-log			linear-log		
K	0.2033	9.74	0.00	75.7670	6.51	0.00
K^2	0.0000	3.16	0.00	0.0003	4.17	0.00
Land	-0.0596	-4.16	0.00	-22.4760	-2.57	0.01
EDU	-1.1840	-3.19	0.00	-939.3500	-3.70	0.00
EDU^2	0.4569	4.32	0.00	305.1200	4.41	0.00
HH	0.0592	1.13	0.26	21.6230	0.83	0.41
DEP	-0.0226	-2.14	0.03	-8.5414	-1.78	0.08
IND	0.0067	7.34	0.00	2.5553	4.93	0.00
Constant	5.6803	13.94	0.00	792.1800	3.03	0.00
Buse R2		0.9994			0.9465	
Log-likelihood value		613.84			-2503.42	
RSS		2.47			572260.00	
	log-linear			linear-linear		
K	0.19	10.60	0.00	80.96	8.08	0.00
K^2	-0.013	-5.16	0.00	-4.5988	-3.43	0.00
LAND	-0.0174	-3.04	0.00	-7.6558	-2.43	0.02
EDU	0.0036	0.09	0.93	-41.3810	-1.68	0.09
EDU^2	0.0052	1.70	0.09	5.1783	2.88	0.00
HH	0.0093	0.87	0.39	6.2726	1.19	0.24
DEP	-0.0226	-2.17	0.03	-8.6538	-1.83	0.07
IND	0.0065	6.87	0.00	2.5312	4.89	0.00
Constant	5.6361	32.95	0.00	424.5800	4.30	0.00
Buse R2		0.9994			0.9469	
Log-likelihood value		617.83			-2496.62	
RSS		2.43			555290.00	

Selection between the two preferred models can be implemented by the χ^2 test derived by Box and Cox (1964), where the null hypothesis is equivalence of the log-linear and linear-linear models, and the test statistic is given by

$$l = \frac{N}{2} \left| \ln\left[\frac{RSS_1 / C^2}{RSS_2} \right] \right|,$$

which is distributed as $\chi^2(1)$, where $C = \exp[(\sum \log Y_i)/N]$, N is the sample size and RSS_1 (RSS_2) denotes the residual sum of squares from the linear-linear (log-linear) models. Substituting the relevant values into the above expression, we obtained l = 73.3. When the null hypothesis is rejected, as in this paper, the log-linear model is selected if $RSS_1/C^2 > RSS_2$. Otherwise, the linear-linear model is selected. Since the ratio of RSS_1/C^2 to RSS_2 is 1.34, the log-linear model is the finally chosen income generating function for rural China.

The log-linear model possesses expected signs for most of the estimated coefficients. After controlling for the dependency ratio, any increase in household size (HH) implies more labour input. From this perspective, the positive sign associated with HH is justified and is consistent with Wan (2004). The negative estimate for the land variable is not unexpected as cropping is known to make loss or little profit (Wan and Cheng 2001). Non-linearity is present for the capital input and possibly the schooling variable. The model fits the data quite well as indicated by the reasonably high R^2 and t-ratios. We decided not to drop the non-significant variables, in order to minimize data-mining.

Before proceeding to inequality decomposition, it is necessary to solve the estimated equation for income Y so inequality is measured over income rather than logarithm of income. Solving the log-linear model, we have:

$$Y = \exp(5.64)\cdot\exp(0.19K - 0.013\,K^2 - 0.017\,\text{Land} + 0.0036\,\text{EDU} + 0.0052\,\text{EDU}^2 + 0.0093\,\text{HH} - 0.0226\,\text{DEP} + 0.0065\,\text{IND} + \text{Loc})\cdot\exp(\text{other terms})$$

where *Loc* is the sum of all regional dummy variable terms and *other terms* is the sum of all year dummy variable terms, time trend and the residual terms. *Loc* represents geographical conditions, weather, water and other non-removable natural resources as well as infrastructure. When relative inequality indicators are used, as in this paper, terms associated with the constant and time trend terms (including year dummy variables) can be removed from the equation without affecting the decomposition results. This is because inequality is measured for each year and these terms are all scalars of the income variable Y (recall the homogeneity theorem of relative inequality measures). Also, the contribution of residual is given by I(Y) – I($\hat{Y}$), where $\hat{Y}$ is the predicted value based on the estimated regression model Thus, the final equation for Shapley decomposition is:

$$\hat{Y} = \exp(0.19K - 0.013\,K^2 - 0.017\,\text{Land} + 0.0036\,\text{EDU} + 0.0052\,\text{EDU}^2 + 0.0093\,\text{HH} - 0.0226\,\text{DEP} + 0.0065\,\text{IND} + \text{Loc})$$

The decomposition is implemented using a Java programme developed by the World Institute for Development Economics Research (UNU-WIDER). Decomposition results

Table 3: Sources of regional inequality in rural China explained by the regression model (%)

	K	Land	EDU	HH	IND	DEP	Loc	Sum	K	Land	EDU	HH	IND	DEP	Loc	Sum
	Gini								Theil-L							
1985	14.98	2.16	18.68	-0.70	15.04	0.95	78.71	129.82	11.06	0.30	23.44	-2.03	17.91	0.86	110.91	162.46
1986	7.93	1.90	16.96	-0.63	13.90	1.47	69.73	111.26	0.12	0.56	18.35	-1.55	15.01	1.59	86.70	120.77
1987	7.03	2.17	15.94	-0.47	14.05	0.50	66.47	105.69	-1.08	0.78	17.05	-1.30	14.92	0.48	81.32	112.10
1988	8.07	2.30	14.38	-0.60	14.04	0.61	64.71	103.51	0.66	0.98	14.53	-1.43	13.98	0.63	76.12	105.44
1989	8.44	2.21	14.26	-0.63	14.63	0.63	64.02	103.56	1.82	0.88	14.58	-1.48	15.69	0.67	76.40	108.55
1990	9.87	2.02	15.90	-0.61	13.25	0.60	63.60	104.62	4.36	0.70	16.60	-1.49	13.25	0.60	74.57	108.63
1991	12.31	1.96	14.73	-0.54	13.73	0.61	61.48	104.28	6.96	0.46	15.24	-1.39	14.47	0.65	73.50	109.89
1992	11.90	2.31	15.88	-0.65	15.02	0.54	59.52	104.51	7.94	0.94	16.78	-1.56	15.83	0.56	70.31	110.83
1993	13.88	2.59	12.51	-0.59	14.59	0.54	54.91	98.43	10.73	1.29	11.86	-1.32	14.04	0.56	60.64	97.84
1994	15.00	2.67	12.69	-0.58	15.27	0.56	52.89	98.51	12.31	1.36	11.81	-1.27	14.55	0.58	57.40	96.74
1995	16.31	1.87	13.10	-0.47	11.60	0.49	47.37	90.27	12.50	0.43	11.80	-0.88	10.11	0.47	47.92	82.36
1996	17.95	2.06	12.16	-0.39	12.10	0.47	45.92	90.26	13.71	0.59	10.88	-0.76	10.67	0.45	46.30	81.84
1997	18.37	2.48	13.02	-0.42	13.17	0.46	46.60	93.69	14.31	1.01	12.31	-0.82	11.33	0.45	47.52	86.11
1998	20.20	2.41	11.76	-0.38	12.54	0.39	46.20	93.12	15.91	0.94	10.54	-0.75	10.75	0.37	46.87	84.62
1999	20.86	2.19	11.67	-0.28	11.32	0.42	45.03	91.21	16.86	0.67	10.44	-0.60	9.38	0.41	45.61	82.79
2000	23.14	2.24	11.26	-0.22	12.28	0.41	44.98	94.09	19.91	0.80	10.55	-0.58	10.67	0.42	46.70	88.47
2001	24.20	2.20	9.43	-0.27	11.06	0.41	44.01	91.05	20.41	0.65	8.01	-0.60	8.89	0.41	44.09	81.88
2002	25.06	2.06	10.93	-0.29	9.45	0.38	43.49	91.08	21.55	0.33	9.80	-0.63	7.01	0.39	43.81	82.25
	Theil-T								CV2							
1985	11.83	0.11	23.56	-2.11	18.33	0.79	111.99	164.55	13.00	-0.24	24.15	-2.22	18.90	0.73	113.84	168.18
1986	-0.83	0.70	18.09	-1.61	15.61	1.61	87.48	121.07	-1.83	0.84	18.05	-1.73	16.35	1.66	88.31	121.70
1987	-2.11	0.92	16.54	-1.34	15.39	0.50	80.97	110.87	-3.24	1.08	16.25	-1.37	15.96	0.54	80.67	109.86
1988	0.00	1.06	14.32	-1.52	14.46	0.64	75.47	104.50	-0.65	1.16	14.21	-1.59	15.04	0.67	74.73	103.58
1989	1.26	0.95	14.17	-1.53	15.95	0.68	75.26	106.75	0.69	1.06	13.92	-1.60	16.30	0.73	74.15	105.25
1990	3.71	0.81	16.26	-1.48	13.71	0.65	74.15	107.81	3.09	0.90	16.03	-1.51	14.24	0.66	73.63	107.06

1991	5.97	0.53	14.57	-1.34	14.57	0.63	71.96	106.88	5.00	0.60	13.99	-1.31	14.70	0.65	70.25	103.88
1992	7.24	1.15	15.95	-1.52	16.23	0.56	69.17	108.82	6.68	1.38	15.22	-1.53	16.75	0.59	68.02	107.13
1993	10.21	1.50	11.23	-1.28	14.34	0.57	59.66	96.23	9.87	1.68	10.58	-1.25	14.66	0.60	58.30	94.43
1994	12.35	1.62	11.37	-1.27	15.24	0.60	57.75	97.64	12.65	1.89	10.91	-1.28	16.07	0.64	58.00	98.89
1995	12.68	0.58	11.40	-0.91	10.56	0.50	48.31	83.12	13.00	0.75	10.92	-0.93	11.13	0.52	48.41	83.80
1996	13.77	0.82	10.49	-0.80	11.33	0.47	46.75	82.84	13.96	1.04	10.10	-0.83	12.09	0.51	46.95	83.83
1997	14.49	1.29	11.83	-0.85	12.13	0.48	48.21	87.60	14.78	1.59	11.37	-0.87	13.05	0.50	48.64	89.05
1998	16.48	1.24	10.23	-0.81	11.71	0.39	47.85	87.12	17.14	1.54	9.88	-0.86	12.79	0.41	48.52	89.44
1999	17.35	0.95	10.01	-0.64	10.16	0.42	46.32	84.56	17.90	1.24	9.56	-0.68	11.08	0.42	46.78	86.32
2000	20.39	1.09	10.16	-0.60	11.54	0.44	47.15	90.15	20.89	1.40	9.80	-0.64	12.56	0.46	47.35	91.81
2001	21.11	0.94	7.61	-0.63	9.64	0.43	44.82	83.91	21.79	1.23	7.18	-0.65	10.47	0.45	45.18	85.66
2002	22.19	0.63	9.25	-0.68	7.54	0.39	44.12	83.45	22.72	0.91	8.69	-0.71	8.16	0.40	43.98	84.14

Source: see text.

are presented in Table 3.[5] Judging from the proportions of inequality explained by the estimated model, our analytical results are quite satisfactory. We can explain 80 per cent or more of total inequality. It is not unexpected that decomposition outcomes vary with inequality measures. However, they are broadly consistent in ranking contributors and in portraying the time trends of individual contributions. Consequently, discussion hereafter will be based on those under Gini only.

For large countries such as China, the variable *Loc* (representing geography, non-removable resources, weather conditions, and so on) is expected to make a very substantial contribution to total regional inequality. This is particularly true for rural China and more so for the early years. In a subsistence society with a closed economy, no market exists and proximity to markets and ports are irrelevant to income. In such cases, a dominant proportion of regional inequality can be explained by the *Loc* variable. In the early days of rural China, markets were fragmented and rural income was almost entirely derived from farming which heavily depends on soil, water and weather conditions. Therefore, it is not surprising to see overwhelming contributions of the *Loc* variable in the 1980s. As non-farming income increases and as markets are developed and infrastructure are improved in poor areas, the percentage contribution of *Loc* started to decline, a finding consistent with a priori expectations[6].

Despite the reduction in its percentage contribution, the importance of *Loc* cannot be over-looked. This is so for several reasons. First, its absolute contribution had been maintained over time. Even in terms of relative contributions, it ranked number one throughout the period under consideration. Second, *Loc* in this paper not only means local natural resources, but also means access to market, information and technology or even investment. The latter directly affects productivity and resource endowment, and indirectly affects efficiency of resource use. Finally, natural resources in terms of land quality and quantity, weather, and water are not subject to market development or infrastructure improvement. It is true that as infrastructure improves for poor regions, locational disadvantages for the inland regions might be alleviated. Nevertheless, these disadvantages could never be eliminated since transportation and communication costs would always be non-decreasing functions of distance to ports and major markets.

Uneven distribution of capital ranks the third in the 1980s but its contribution gradually increased to over 25 percent, making it the second largest contributor to regional inequality in rural China after 1996. This result is in line with Wan and Zhou (2005)

---

[5] The results are different from Wan (2004), who did not use the regional price deflators of Brandt and Carsten (2004). A different modelling strategy was also followed in Wan (2004), where location was not fully accounted for.

[6] The decomposition results for 1985 and 1986 seem to produce unusually large contributions of the residual term. But our results look more acceptable than those of Morduch and Sicular (2002) and they are consistent over years in terms of trends of the individual contributions.

who used household data instead of aggregate regional data. On the other hand, education was the second largest contributor and its position more or less matched that of industrialization since 1996. The contribution of industrialization is large but smaller than that suggested by earlier studies of Rozelle (1994) and Wan (2001). Such an inconsistency is most likely due to contamination in early analytical frameworks, where other factors were not controlled for. Wan (2004) obtained a smaller contribution than Rozelle (1994) and Wan (2001) after controlling for some variables. However, Wan (2004) did not incorporate regional dummy variables in his income function. When location dummies are included, as in this study, the contribution of industrialization is bound to become even smaller.

The only negative contributor is household size. As it represents labour input and labour is more abundant in poor regions, such a finding is justified. In reality, the household size may imply extra income from side line activities (Wan 2004). Unfortunately, this only equalizing factor only makes negligible impacts on the total inequality. Furthermore, as household size converges in China, this mild equalizing contribution will disappear in the long run. Related to the household size variable, dependency ratio makes a negligible but positive contribution. Land is found to be an inequality-increasing contributor. This is caused by the negative returns to land in China for many of the years under study. It is known that poor regions possess more land and poor households are those who mainly engage in farming. If government support is effective enough to reverse the sign of the land marginal product, land would be an equalizing factor in rural China.

5 Summary and policy implications

In this paper, we constructed a time profile of rural regional inequality in China over the period 1985-2002. We further decomposed total regional inequality into between (east-central-west) belts contribution and contributions due to regional income gaps within these belts. Both the conventional decomposition and the proposed Shapley value decomposition yielded similar results. Finally, we applied the inequality accounting framework of Wan (2004) to identify root sources of the total inequality. Several findings deserve special mention.

First, regional income is found to diverge, more between regional belts than within these belts. In other words, inequality between regional belts as well as that within these belts has been on the rise. The fast increase in regional inequality is accompanied by worsening polarization. Second, while the east-central-west divide constituted some 50 per cent of the total regional inequality in the mid 1980s, its contribution increased to around 60 per cent as from 1996. Third, location and location-related factors comprise the largest contributor to total regional inequality although its percentage contribution has decreased over time. Fourth, capital and rural industrialization are the second and third largest contributors to total inequality. Finally, schooling or human capital has been gaining importance as a determinant of regional inequality. Based on these findings, we can derive the following policy implications:

(a) National policy must target regional belts, not only individual regions. As farming structure becomes more homogenous in neighbouring regions, policy-induced and other shocks are likely to enhance polarization unless supplementary measures are taken at the stage of policy design.

(b) While infrastructure investment in inland regions are necessary, more attention should be given to capital accumulation at the household level in the poor regions. It is possible that capital accumulation may become the largest contributor to regional inequality in the not too distant future. Thus, development of rural capital market, particularly credit access to the poor, should be placed on the top agenda of central and local governments.

(c) More concerted efforts must be devoted to human capital accumulation in poor areas. Schooling might not matter so much in largely subsistence China in the 1970s or early 1980s, such is not the case any more. The growing contribution of schooling to regional inequality appeals for serious government educational input in the interior regions.

(d) Continued support for generating non-farming incomes in the poor regions can lead to substantial reduction in regional inequality. Fiscal and budgetary policies should make allowance for the initiation and growth of rural industries in the inland regions as far as inequality reduction is concerned.

(e) Finally, much more is needed than the abolition of agriculture tax in providing assistance to grain farmers. This is not only has bearing on the food security of the nation, but also is potentially effective for combating the high level of regional inequality in China.

Appendix

Table A1: Composition of regional inequality in rural China

Year	Between belts	Within belt of east	central	west	Total Theil-L
1985	0.010	0.002	0.004	0.004	0.020
1986	0.012	0.003	0.006	0.005	0.025
1987	0.015	0.003	0.005	0.004	0.027
1988	0.015	0.004	0.005	0.005	0.029
1989	0.018	0.003	0.004	0.004	0.030
1990	0.018	0.004	0.006	0.004	0.032
1991	0.019	0.004	0.005	0.004	0.032
1992	0.023	0.005	0.005	0.004	0.036
1993	0.029	0.006	0.004	0.004	0.043
1994	0.030	0.005	0.007	0.004	0.046
1995	0.040	0.005	0.007	0.004	0.056
1996	0.039	0.007	0.007	0.005	0.058
1997	0.038	0.008	0.006	0.005	0.057
1998	0.036	0.008	0.006	0.006	0.056
1999	0.037	0.008	0.006	0.006	0.058
2000	0.037	0.008	0.006	0.006	0.057
2001	0.041	0.008	0.006	0.006	0.061
2002	0.041	0.008	0.006	0.007	0.062

Table A2: Shapley decomposition of regional inequality into between- and within- components

	Between	Within belt of				Between	Within belt of			
Year	belts	east	central	West	Total	belts	east	central	west	Total
			Gini					Theil-L		
1985	0.050	0.019	0.024	0.016	0.109	0.010	0.002	0.004	0.003	0.020
1986	0.054	0.023	0.029	0.018	0.123	0.013	0.003	0.005	0.004	0.025
1987	0.063	0.025	0.025	0.017	0.129	0.015	0.004	0.005	0.003	0.027
1988	0.060	0.029	0.025	0.019	0.134	0.015	0.005	0.005	0.004	0.029
1989	0.070	0.027	0.022	0.018	0.137	0.018	0.004	0.004	0.004	0.030
1990	0.069	0.028	0.027	0.017	0.141	0.018	0.005	0.005	0.003	0.032
1991	0.074	0.028	0.024	0.017	0.142	0.020	0.005	0.004	0.003	0.032
1992	0.079	0.032	0.025	0.015	0.151	0.022	0.006	0.004	0.003	0.036
1993	0.094	0.035	0.022	0.013	0.164	0.029	0.008	0.004	0.003	0.043
1994	0.092	0.034	0.030	0.014	0.170	0.030	0.007	0.007	0.003	0.046
1995	0.110	0.034	0.030	0.012	0.186	0.039	0.007	0.006	0.003	0.056
1996	0.104	0.039	0.032	0.013	0.188	0.038	0.009	0.007	0.003	0.058
1997	0.104	0.041	0.027	0.014	0.186	0.037	0.011	0.006	0.004	0.057
1998	0.100	0.042	0.028	0.016	0.186	0.035	0.011	0.006	0.004	0.056
1999	0.102	0.043	0.028	0.016	0.188	0.036	0.011	0.006	0.005	0.058
2000	0.103	0.042	0.027	0.016	0.188	0.036	0.011	0.006	0.004	0.057
2001	0.108	0.042	0.028	0.016	0.195	0.039	0.011	0.006	0.005	0.061
2002	0.110	0.042	0.027	0.016	0.195	0.040	0.011	0.006	0.005	0.062
			Theil-T					CV^2		
1985	0.010	0.002	0.004	0.003	0.019	0.019	0.005	0.008	0.005	0.037
1986	0.012	0.003	0.005	0.003	0.024	0.024	0.008	0.010	0.006	0.047
1987	0.015	0.004	0.004	0.003	0.026	0.030	0.009	0.009	0.005	0.052
1988	0.015	0.005	0.005	0.004	0.028	0.030	0.012	0.009	0.006	0.057
1989	0.018	0.005	0.004	0.003	0.029	0.036	0.010	0.007	0.005	0.059

1990	0.018	0.005	0.005	0.003	0.031	0.036	0.012	0.009	0.005	0.062
1991	0.019	0.006	0.004	0.003	0.032	0.040	0.013	0.007	0.005	0.065
1992	0.022	0.007	0.004	0.003	0.036	0.045	0.015	0.008	0.004	0.072
1993	0.028	0.008	0.003	0.002	0.042	0.057	0.019	0.007	0.004	0.086
1994	0.029	0.008	0.006	0.003	0.045	0.057	0.017	0.012	0.004	0.090
1995	0.037	0.008	0.006	0.002	0.054	0.074	0.018	0.012	0.003	0.107
1996	0.036	0.010	0.007	0.003	0.055	0.069	0.021	0.014	0.004	0.109
1997	0.035	0.011	0.006	0.003	0.054	0.069	0.023	0.011	0.004	0.106
1998	0.033	0.011	0.006	0.003	0.053	0.065	0.023	0.011	0.005	0.105
1999	0.035	0.011	0.006	0.003	0.055	0.068	0.024	0.011	0.006	0.109
2000	0.035	0.011	0.006	0.003	0.055	0.069	0.024	0.011	0.005	0.110
2001	0.038	0.012	0.006	0.003	0.059	0.075	0.026	0.012	0.005	0.118
2002	0.039	0.011	0.006	0.003	0.059	0.077	0.025	0.011	0.005	0.119

References

Box, G.E.O., and Cox, D.R. (1964). 'An Analysis of Transformation', *Journal of the Royal Statistical Society* A26: 211-52.

Brandt, L., and A.H. Carsten (2004). 'Spatial Price Differences in China: Estimates and Implications' (mimeo), Social Science Division, University of Science and Technology: Hong Kong.

Cancian, M., and D. Reed (1998). 'Assessing the Effects of Wives Earning on Family Income Inequality', *Review of Economics and Statistics* 80: 73-9.

Dagnum, C. (1990) 'On the Relationship between Income and Inequality Measures and Social Welfare Functions', *Journal of Econometrics* 43: 91-102.

Fields, G.S. (2001). *Distribution and Development: A New Look at the Developing World*, MIT Press: Cambridge MA.

Fields, G.S., and G. Yoo (2000). 'Falling Labour Income Inequality in Korea's Economic Growth: Patterns and Underlying Causes', *Review of Income and Wealth* 46: 139-59.

Kanbur, R. (2002). 'Notes on the Policy Significance of Inequality Decompositions' (mimeo), Cornell University: Ithaca NY.

Kanbur, R., and X. Zhang (1999). 'Which Regional Inequality: Rural–Urban or Coast–Inland? An Application to China', *Journal of Comparative Economics* 27: 686-701.

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

Kmenta, J. (1986) *Elements of Econometrics*, Prentice-Hall: New Jersey.

Morduch, J., and T. Sicular (2002). 'Rethinking inequality decomposition, with evidence from rural China', *The Economic Journal* 112: 93-106.

NBS (National Bureau of Statistics) (various). *China Rural Statistical Yearbook*. China Statistical Publishing House: Beijing.

NBS (National Bureau of Statistics) (various). *China Rural Household Survey Yearbook*, China Statistical Publishing House: Beijing.

Ravallion, M., and S. Chen (2004). China's (Uneven) Progress Against Poverty, *Policy Research Working Paper* 3408, World Bank: Washington DC.

Rozelle, S. (1994). 'Rural Industrialization and Increasing Inequality: Emerging Patterns in China's Reforming Economy', *Journal of Comparative Economics* 19: 362-91.

Shorrocks, A.F. (1984). ‘Inequality Decomposition by Population Subgroups’, *Econometrica* 52(6): 1369-85.

Shorrocks, A.F. (1999). ‘Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapely Value’, (unpublished manuscript) Department of Economics, University of Essex: Colchester.

Shorrocks, A., and D. Slottje (2002). ‘Approximating Unanimity Orderings: An Application to Lorenz Dominance’, *Journal of Economics* (Supplement 9): 91-118.

Shorrocks, A.F., and G.H. Wan (2005). ‘Spatial Decomposition of Inequality’, *Journal of Economic Geography* 5(1): 59-82.

Sicular, T., X. Yue, B. Gustafsson, and S. Li (2007). ‘The Urban–Rural Gap and Income Inequality in China’, forthcoming in *Review of Income and Wealth*.

Silber, J. (1989). ‘Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality’, *Review of Economics and Statistics* 71(1): 107-15.

Tsui, K.Y. (1991). ‘China's Regional Inequality, 1952-85’, *Journal of Comparative Economics* 15: 1-21.

Wan, G. (2001). Changes in Regional Inequality in Rural China: Decomposing the Gini Index by Income Sources, *Australian Journal of Agricultural and Resource Economics* 43: 361-81.

Wan, G. (2002). Regression-based Inequality Decomposition: Pitfalls and a Solution Procedure, *WIDER Discussion Paper* 2002/101, UNU-WIDER: Helsinki.

Wan, G.H. (2004), ‘Accounting for Income Inequality in Rural China’, *Journal of Comparative Economics* 32(2): 348-63.

Wan G.H. (2005). ‘Rising Inequality in Post Reform China’, *WIDER Angle* 2/2005, UNU-WIDER: Helsinki.

Wan, G.H., and E.J. Cheng (2001). ‘Effects of Land Fragmentation and Returns to Scale in the Chinese Farming Sector’, *Applied Economics* 33(2): 183-194.

Wan, G.H., M. Lu, and Z. Chen. (2006). ‘The Inequality-Growth Nexus in the Short- and Long- run: Empirical Evidence from China’, *Journal of Comparative Economics* 34(4): 654-7.

Wan, G.H., and Z. Zhou (2005). ‘Income Inequality in Rural China: Regression-Based Decomposition Using Household Data’, *Review of Development Economics* 9: 107-20.

Zhang, X. (2005). ‘Sang Nong Problems and Rural Development’, online paper, http://www.ccrs.org.cn/article_view.asp?id=108

Zhang, Y., and G.H. Wan (2006), ‘The Impact of Growth and Inequality on Rural Poverty in China’, *Journal of Comparative Economics* 34(4): 694-712.