United Nations University
WIDER
World Institute for Development Economics Research

Research Paper No. 2004/51

Income Inequality in Rural China

Regression-Based Decomposition Using Household Data

Guanghua Wan[1] and Zhangyue Zhou[2]

August 2004

Abstract

A considerable literature exists on the measurement of income inequality in China and its increasing trend. Much less is known, however, about the driving forces of this trend and their quantitative contributions. Conventional decompositions, by factor components or by population subgroups, only provide limited information on the determinants of income inequality. This paper represents an early attempt to apply the regression-based decomposition framework to the study of inequality accounting in rural China, using household level data. It is found that geography has been the dominant factor but is becoming less important in explaining total inequality. Capital input emerges as a most significant determinant of income inequality. Farming structure is more important than labour and other inputs in contributing to income inequality across households.

Keywords: inequality decomposition, regression, income generating function, China
JEL classification: D33, R12, O53

---

[1]UNU-WIDER, Helsinki; [2]Asian Agribusiness Research Centre, University of Sydney

This study has been prepared within the UNU-WIDER project on Spatial Disparities in Human Development, directed by Ravi Kanbur and Tony Venables, with Guanghua Wan.

UNU-WIDER acknowledges the financial contributions to the 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 92-9190-640-9 (internet version)

Acknowledgements

We are grateful to the referees, and to Yin Zhang for useful comments on early versions of this paper.

UNU World Institute for Development Economics Research (UNU-WIDER)

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Camera-ready typescript prepared by Lorraine Telfer-Taivainen at UNU-WIDER

Printed at UNU-WIDER, Helsinki

1 Introduction

Many studies have appeared in both English and Chinese, focusing on income distribution in rural China.[1] They point to a worsening trend since the late 1970s when China initiated economic reforms. Such a trend has serious implications on China's ability to maintain sustainable growth and, if unabated, will undermine social and political stability. The issue of income distribution ranks as one of the top priorities in the government's policy agenda. The two most important national conferences held recently in March 2003—the National People's Congress and the Chinese People's Political Consultative Conference—expressed unprecedented concerns about rural income and income inequality.

While a consensus has been reached about the increasing trend in income inequality in rural and urban China, this is not the case regarding the causes of such increases. Generally speaking, variables affecting income generation will also determine income inequality. Thus, economic theory and common knowledge can be used to identify these variables. In other words, one could easily compile a list of factors which may explain income gaps, such as different resource endowments and policy biases. However, for the purpose of setting policy priorities, it is necessary to rank the variables in terms of their relative contributions to total inequality. This usually requires inequality decomposition.

Conventional approaches to inequality decomposition typically follow Shorrocks (1980, 1982, 1984) and Bourguignon (1979). Under these frameworks, one can carry out decomposition either by population subgroups or by factor components. The former produces the so-called 'within' and 'between' components. It has been used to examine issues such as urban–rural income gaps, male–female wage differentials, and so on. For a recent reference, see Shorrocks and Wan (2004a). For example, Kanbur and Zhang (1999) find that regional income inequality in China consists of 70-78 percent of between (urban and rural) component and the remaining is within component. This kind of decomposition is silent on the fundamental determinants of either of the two components. Also, the decomposition is likely to produce spurious results. For example, decomposition of wage inequality by gender might give rise to a sizable between-gender component. However, this may have little to do with sex discrimination in the work place if females are less educated before they enter the labour market, a phenomena not uncommon in many developing countries. Similarly, a large between-race component may have little to do with skin colour unless other personal attributes such as education, age, occupation and so on can be assumed to be identical. Clearly, one must be able to control for other factors in order to identify and measure the contribution of a particular variable. This is not possible with the conventional approaches.

Decomposition by factor components requires complete information on all income sources. It also requires an identity that expresses total income as a sum of factor incomes. Apart from a data unavailability problem, this approach cannot be used to quantify contributions of fundamental determinants to income inequality either. For example, it is known that

[1] See Wan (2001) and references therein.

income is determined by education, experience and other personal or household characteristics. These fundamental determinants affect all sources of income, including wage, investment returns and transfer income. It would be interesting and useful to decompose total inequality into components associated with each of the fundamental determinants. However, decomposition by factor components only allows one to attribute total inequality to the income sources, not to the fundamental determinants.

This paper contributes to the literature on income inequality in rural China in a number of ways. It represents an early attempt to analytically identify the fundamental determinants of income inequality in rural China. The use of regression-based decomposition is novel in that it allows ranking of these determinants according to any inequality measure. Moreover, household-level data are used in this paper which complements the existing literature mostly based on aggregate data.

In the next section, we present a brief discussion on income disparity in rural China and on data source. Section 3 describes the regression-based decomposition technique and the income generation function. This is followed by interpretation of decomposition results and policy implications in Section 4. Section 5 concludes.

2 Income disparity in rural China and the data source

Income disparity can be examined at different levels of aggregation. At the national level, provinces or regions (in some cases, representative counties) are usually taken as the unit of analysis. This is the basis of most publications on rural income inequality in China.[2] In this context, income gaps are found to be large. For example, in 2002 per capita annual net income in rural Shanghai was 6,224 yuan while that in rural Guizhou was only 1,490 yuan. As shown in Table 1, rural incomes are generally higher in the relatively developed east. Most provinces located in central China possess per capita rural incomes around the national average. All those with a per capita income below 2,000 yuan are located in the west.

Over the years, the interregion income gaps have risen. In 1985, the highest per capita rural net income was 3.2 times that of the lowest. This ratio has increased to 4.3 in 2002 (SSB 2003:368). When provinces are ranked in terms of per capita income level, the rankings change little from year to year. This is particularly true regarding the top and bottom positions. This suggests that convergence has not taken place in China despite continuous economics growth at the national and regional levels.

Based on data from three provinces with different development levels, Figure 1 indicates that the income gaps were relatively small in the mid 1980s but expanded rapidly in the

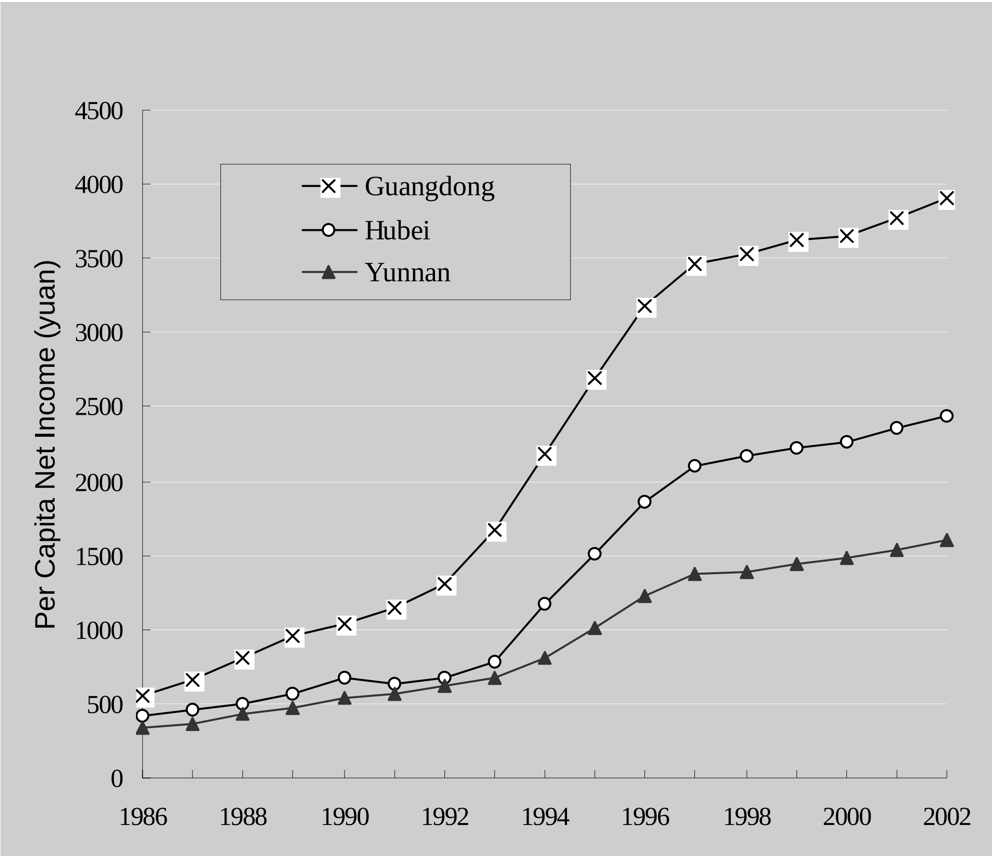
[2] For a comprehensive discussion on inequality in China (not only rural China), see Kanbur and Zhang (2005).

Table 1: Per capita income in rural China, by province (2002)

Province	Yuan	Province	Yuan
Shanghai	6224	Henan	2216
Beijing	5398	Shanxi	2150
Zhejiang	4940	Anhui	2118
Tianjin	4279	Sichuan	2108
Jiangsu	3980	Chongqing	2098
Guangdong	3912	Inner Mongolia	2086
Fujian	3539	Guangxi	2013
Shandong	2948	Ningxia	1917
Liaoning	2751	Xinjiang	1863
Hebei	2685	Qinghai	1669
Hubei	2444	Yunnan	1609
Hainan	2423	Shaanxi	1596
Heilongjiang	2405	Gansu	1590
Hunan	2398	Guizhou	1490
Jiangxi	2306	Tibet	1462
Jilin	2301	national average	2476

Source: SSB (2003:368).

Figure 1: Per capita rural income, selected provinces



Source: SSB (various).

mid 1990s.[3] This expansion implies that factors other than geography must have played a more and more important role. Figure 1 also shows that income growth in rural China has slowed down significantly since the mid 1990s. Given China's consistent growth performance, the slow-down implies a worsening urban–rural gap. The slower growth, coupled with rising inequality, naturally hinders progress in poverty eradication.

Not only is there a wide income disparity between provinces, significant income inequality also exists among villages within a province and among households within a village. Table 2 reports frequency distribution of household income for nine villages. The last row of the table reveals that in Yunan province, village 1 possesses a level of per capita income that is 12 times that of village 2 in the same province. Within Hubei, 65 percent of households in village 3 have per capita incomes below 2,000 yuan while this percentage is only 17 in village 2. Income difference is also evident across villages in Guangdong. As explored later, the inter-village income differences accounts for some 40 percent of total inequality. At the household level, the gap is even larger. In Yunan 55 percent of households in village 2 have a per capita income below 500 yuan while in Guangdong, over 80 percent of households in village 1 have a per capita income over 10000 yuan. Such large inter-household income gaps imply an alarming level of inequality in rural China.

The high and rapidly rising inequality in China has attracted considerable attention.[4] In what follows, we will use household-level survey data to compute various inequality indices and conduct inequality decomposition. The data are collected by the Research Centre for Rural Economy (RCRE) of the Ministry of Agriculture of China. The RCRE survey began in 1986, and has since been conducted every year except for 1992 and 1994. All households covered by the survey are asked to keep records of incomes and expenses, and other information. These are collected, checked, processed and reported by the survey team. The survey instruments have evolved over the years. Those used for 1986-91 were the same (with 312 variables). They were expanded for the 1993 survey (with 394 variables) and further expanded in 1995 (with 439 variables). Data between 1995 and 2002 are used in this study to ensure the consistency of variables over time.

It is not possible to access the complete data set. For this study, we use data from three provinces, Guangdong, Hubei and Yunnan. Guangdong, located in south-east China, is among the richest provinces. Hubei, a province in central China, is of a medium development status. Western China is represented by Yunnan, a well-known poor province. From each province, three villages are chosen, representing different development status within the county (see Table 2). While not claiming to be

---

3 Guangdong represents an economically developed region, Hubei a medium developed region, and Yunnan an economically less developed region.

4 See, for example, Griffin and Zhao (1993), Rozelle (1996), Hu et al. (1997), Yao (1997), Ravallion and Chen (1999), and Wu (2000:261-81). Except Kanbur and Zhang (1999, 2005), most of the published works only provide a snapshot, without a time profile. Many of them use proxy variables such as agricultural output (Howes and Hussain 1994), regional national income (Tsui 1991), collective income (Griffin and Saith 1982) or even grain output (Lyons 1991), rather than personal income. These proxies may not adequately represent living standards in China (Wei et al. 1997). These deficiencies are recognised by Tsui (1991), Knight and Song (1993), Chen and Fleisher (1996). Chen and Fleisher (1996) explicitly appeal for the use of per capita income data to address the inequality issue in China.

representative of China, the data do cover a variety of geoeconomic conditions and are more representative than studies relying on data from a single province or single county. Notwithstanding the novelties discussed earlier, this paper can be viewed as an extension to Morduch and Sicular. Morduch and Sicular (2002) use survey data of 259 rural households in Zouping county of Shandong province, covering the period of 1990-93.

3 Regression-based decomposition and income generation function

The regression-based decomposition methodology was proposed in the early 1970s (Blinder 1973; Oaxaca 1973) but had not gained much attention until recently (see Juhn et al. 1993; Bourguignon et al. 2001). Wan (2002) provides a detailed account on the development of this technique. For recent empirical applications, see Fields and Yoo (2000), Adams (2002), Morduch and Sicular (2002), Heltberg (2003), Zhang and Zhang (2003), and Wan (2004).

As the first step of the regression-based decomposition, income generation function must be obtained. In specifying such a function for rural China, consideration must be given to both human capital theory and production theory. This is because farmers, unlike wage earners, must use land and physical capital in addition to labour in deriving their income. Thus, standard production inputs of land, labour and capital should be included. The human capital theory calls for inclusion of skill variables such as education, training and experience (often represented by age). As an accepted practice in the development literature, the education level and age of the household head will be used.

It is also necessary to consider factors which could alter income even if production inputs and human capital are the same. One such factor is the type of business activity that a household engages in, by which households are classified into 10 different categories by the RCRE. These include cropping, forestry, animal husbandry, fishery, industry, construction, transportation, retailing, food and other services, and finally no business activity. These indicate the main sector from which a household derives most of its income. Clearly, a set of dummy variables is needed to capture differences in income levels arising from different business activities. These dummy variables, taken together, will be referred to as a sector indicator. On the other hand, it is known that grain-cropping in China is often enforced administratively due to low or negative returns (Wan 2004).

Consequently, two identical households may receive different income simply because one grows grain and the other grows vegetable or other cash crops. Thus, the cropping pattern is crucial, which is defined as the ratio of area sown to grain crops over total sown area. Finally, consider two rural households with the same amount of resources but one with wage earners and the other not. Wage earners are those working for the government or industries not run by the household. The number of wage earners reflects the level of urbanization, thus its inclusion in the model enables one to make inference about the impact of urbanization on income inequality in rural China. Ideally, urbanization should be defined at the town or county level. However, this is not possible given the availability of household-level data only.

Table 2: Frequency distribution of rural household income in selected provinces in China, 2002

Per Capita Income	Hubei			Guangdong			Yunnan			All
(yuan)	village 1	village 2	village 3	village 1	village 2	village 3	village 1	village 2	village 3	villages
<500			10		1			55		8
500 –1000	7		20		10	9		33	12	10
1000-2000	20	17	35		32	36	1	11	45	22
2000-3000	28	47	22	1	27	33	8	1	27	20
3000-4000	8	22	2	3	14	16	12		9	9
4000-6000	23	13	10	1	7	6	34		6	11
6000-8000	3		2	4	4		22		1	4
8000-10000	5	2		4	2		5			2
>10000	5			87	2		17			14
Number of households	60	60	60	100	97	100	99	100	100	776
Village-level per capita income	4124	2875	1880	22519	2801	2255	7076	578	2101	5477

Source: Authors' calculations based on RCRE household-level survey data.

6

Geography is important in income determination as it is closely related to non-removable resources as well as to market access, infrastructure and local culture. Data unavailability prevents direct inclusion of geographic variables. However, given the control for physical and human capital inputs and other factors, village dummies can be used to capture the effects of geography or location.It is noted that inclusion of these village dummies does not necessarily entail a fixed-effects model as household-level observations are to be used to estimate the income generation function. Finally, year dummies are included in the estimation to take into account technical changes and reform impacts. The variables included in the income function are given below.[5]

Dependent variable:

Income: per capita annual net income.

Independent variables (dummy variables not listed):

Capital: per capita capital stock;
Land: per capita arable land area;
Labour: number of labourers divided by household size;
Wage earner: proportion of wage earners in household labour force;
Education: number of schooling years of household head;
Education squared;
Training: proportion of household members who received vocational training;
Age: age of household head;
Age squared;
Grain: ratio of grain sown area to total sown area.

The choice of the parametric functional form is dictated by the standard Mincer model, augmented with production inputs and other variables. In other words, the income generation function takes the form of:

Ln (Income) = f (Land, Labour, Capital, …, dummy variables),

where f stands for the standard linear function. The use of the semilog specification is also prompted by the finding that the income variable can be approximated well by a lognormal distribution (Shorrocks and Wan 2004b). The panel data model can be estimated by various techniques. However, the iterative GLS method outlined in Kmenta (1986) is found to work well with Chinese data (Wan and Cheng 2001). This method allows for both heteroscedasticity across households and autocorrelation over time. The model estimation results are tabulated in Table 3.

Leaving the dummy variables aside, all coefficient estimates are of the expected signs and most of them are statistically significant at the 1 percent or 5 percent level of significance. In particular, the negative estimates for the quadratic age and quadratic education variables

5 Morduch and Sicular (2002) also consider political variables in their model, represented by membership of the Communist Party and presence of government officials in households. However, they find that these variables contribute little to income generation or inequality. Surprisingly, Morduch and Sicular (2002) do not include the capital variable, which is a most important contributor to income generation and inequality in China (Zhang and Zhang 2003).

are consistent with standard human capital theory. As expected, the cropping pattern variable, denoted by ‘Grain’ in Table 3, has a negative and significant coefficient estimate.

Table 3: Estimated income generation function (dummy variables not included)

Variable	Coefficient estimate	T-ratio	Level of significance
Capital	0.0958	15.59	0.000
Land	0.0192	2.59	0.009
Labor	0.5999	17.18	0.000
Wage earner	0.0224	3.43	0.001
Education	0.1365	3.72	0.000
Education^2	-0.0107	-1.51	0.130
Training	0.1318	2.74	0.006
Age	0.1450	4.88	0.000
Age squared	-0.0255	-5.33	0.000
Grain	-0.3164	-11.72	0.000
Constant	7.0841	84.61	0.000

Note: Loglikelihood value = -4648.32. Sample size = 6121.

Source: Authors’ calculations.

For a given income generation function, alternative approaches can be used to decompose total income inequality (Wan 2002). Note, however, that the semilog specification implies a nonlinear income generation function in terms of the original income variable. Thus, the Shapley value framework of Shorrocks (1999) must be adopted in a regression-based decomposition context. The constant term becomes a scalar once the estimated semilog function is solved for the original income. It can be ignored in inequality measurement or decomposition as long as relative inequality measures are used. The same can be said to the yearly dummy variables which differentiate income generation functions for different years, only by differences in the constant term.

The Shapley value decomposition involves rather extensive computing. Suppose $Y = f(X_1,\dots, X_K)$ is a general income generation function. Usually Xs are different for different individuals.Replacing X_k by its sample mean would eliminate any differences in X_k among individuals.It is easy to re-compute Y after this replacement. The resulting income, denoted by Y_k, differs from individual to individual because Xs other than X_k differ for different individuals. However, the differences cannot be attributed to X_k any more. In other words, inequality in Y_k, denoted by $I(Y_k)$ is due to differences in Xs excluding X_k. According to the most natural rule of Shorrocks (1999), the contribution of X_k to total inequality, C_k, can be obtained as $I(Y) – I(Y_k)$ for $k = 1, \dots, K$. Shorrocks (1999) terms these contributions the first round effect, which is obtained when only one independent variable X_k is replaced by its sample mean. One can obtain a second round C_k by replacing two variables X_k and X_j with their sample means in computing Y_{kj}. The second round contribution can be written as $C_k = I(Y_j) – I(Y_{jk})$ for $k, j = 1, \dots, K$ $(k \neq j)$. By the same token, the third round contribution can be obtained as $C_k = I(Y_{ij}) - I(Y_{ijk})$ for $k, j, i = 1, \dots, K$ $(k \neq j \neq i)$. This process continues until all Xs are replaced by their sample means. At each round, it is possible to have multiple C_k, which are averaged first and then averaged across all rounds—see Shorrocks (1999) for details.

What about the residual term? Admittedly, one may not be able to analyse the residual contribution. However, it can show how much the estimated model explains total inequality. If the model only explains 30-40 percent of total inequality, leaving the rest to the residual term, policy-makers may well be advised not to rely on the decomposition results. In this study, the residual term is dealt with according to the procedure proposed in Wan (2002, 2004). With the semilog income generation function, the contribution of the residual term can be easily computed as the difference between total inequality and the sum of contributions of all explanatory variables.

4 Decomposition results and discussions

Table 4 tabulates total inequality by various measures. It is clear that the CV^2 indicates a small dip in 1999 and a substantial reduction in 2001. Other measures indicate two slight dips in 1998 and 2001. Nevertheless, they all point to an increasing trend. Since these inequality values are obtained using household-level data, they must be larger than those based on aggregate data. Use of provincial or county-level data only permits measurement of the between province or between county component while what are presented in Table 4 contain all within components (within province, within county and within village).

Table 4: Total income inequality, 1995-2002

	Gini	Akinson	Theil-L	Theil-T	CV2
1995	0.467	0.322	0.388	0.403	1.282
1996	0.505	0.370	0.462	0.482	1.667
1997	0.509	0.371	0.464	0.548	3.006
1998	0.500	0.358	0.443	0.541	3.259
1999	0.520	0.399	0.509	0.567	3.122
2000	0.553	0.433	0.567	0.684	4.547
2001	0.537	0.419	0.543	0.592	2.664
2002	0.638	0.539	0.774	0.907	5.761

Source: Authors' calculations.

Table 5 presents the decomposition results for selected years, where inequality is measured by two different indicators.[6] Both absolute and percentage contributions are shown in the table. Not unexpected, different measures give rise to different decomposition results. This is because different measures are underlined by different social welfare functions and are sensitive to different segments of the Lorenz curve. Nevertheless, some broadly consistent findings can be drawn from Table 5. In reality, one has to choose a particular inequality measure when inconsistent results are obtained. Due to its popularity, the Gini values will be used for discussions hereafter.

6 CV^2 is known to be inferior to other inequality measures as it violates the transfer axiom. Further, results under Theil-T and CV^2 are similar. The Atkinson index is a monotonic transformation of the Theil-L. Thus, we prefer to use the Gini and the Theil-L.

Table 5: Decomposition results, selected years

	Gini	%	Thiel-L	%	Year
Capital	0.0113	2.23	0.0085	1.84	
Land	-0.0018	-0.36	-0.0053	-1.15	
Labour	0.0259	5.13	0.0086	1.87	
Wage earner	0.0102	2.02	0.0076	1.64	
Education	0.0170	3.36	0.0110	2.37	
Training	0.0039	0.77	0.0022	0.48	1996
Age	0.0051	1.01	0.0017	0.38	
Grain	0.0407	8.06	0.0287	6.22	
Sector dummies	0.0384	7.61	0.0227	4.91	
Village dummies	0.2545	50.43	0.2105	45.55	
All Xs	0.4052	80.27	0.2963	64.11	
Total	**0.5048**	**100**	**0.4621**	**100**	
Capital	0.0182	3.64	0.0150	3.39	
Land	-0.0021	-0.42	-0.0055	-1.25	
Labour	0.0233	4.66	0.0059	1.34	
Wage earner	0.0107	2.15	0.0082	1.85	
Education	0.0173	3.45	0.0116	2.61	
Training	0.0036	0.71	0.0021	0.48	1998
Age	0.0051	1.03	0.0019	0.42	
Grain	0.0452	9.03	0.0336	7.59	
Sector dummies	0.0348	6.96	0.0216	4.88	
Village dummies	0.2600	51.98	0.2161	48.82	
All Xs	0.4162	83.19	0.3104	70.13	
Total	**0.5003**	**100**	**0.4427**	**100**	
Capital	0.0885	16.00	0.1112	19.61	
Land	-0.0022	-0.41	-0.0060	-1.06	
Labour	0.0271	4.91	0.0105	1.84	
Wage earner	0.0106	1.92	0.0089	1.56	
Education	0.0154	2.78	0.0096	1.69	
Training	0.0024	0.43	0.0013	0.22	2000
Age	0.0045	0.82	0.0012	0.21	
Grain	0.0486	8.79	0.0419	7.39	
Sector dummies	0.0451	8.15	0.0402	7.08	
Village dummies	0.2591	46.85	0.2366	41.71	
All Xs	0.4990	90.23	0.4552	80.26	
Total	**0.5531**	**100**	**0.5672**	**100**	
Capital	0.1517	23.76	0.2106	27.20	
Land	-0.0026	-0.40	-0.0066	-0.85	
Labour	0.0239	3.75	0.0086	1.11	

table continues…

Wage earner	0.0100	1.56	0.0085	1.10	
Education	0.0132	2.07	0.0071	0.92	
Training	0.0057	0.90	0.0084	1.08	2002
Age	0.0045	0.70	0.0010	0.13	
Grain	0.0494	7.73	0.0476	6.14	
Sector dummies	0.0551	8.63	0.0589	7.60	
Village dummies	0.2544	39.84	0.2547	32.89	
All Xs	0.5653	88.54	0.5988	77.32	
Total	**0.6384**	**100**	**0.7744**	**100**	

Source: Authors' calculations.

Referring to Table 5, geography as represented by village dummies contributes a fairly constant amount to total inequality. This finding reflects the fact that spatial distribution of geographic factors cannot be easily altered in the short or medium run. Since total inequality has been increasing over time, the percentage contribution of geography displays a decreasing trend. Despite this, geography still explains almost 40 percent of the total inequality in 2002, 15 percent lower than in 1995. Apart from its role in determining market access, geography is closely associated with natural resource endowments such as water and weather conditions. Natural resources are particularly crucial for farm production activities and they are neither tradable nor removable. Infrastructure provision may improve market access for the poor areas, but it could also benefit the rich regions. Thus, the overall impact of infrastructure development on total inequality could be small. Needless to say, geography will continue to play an important role in constituting rural income inequality. This finding can be used to justify regional development policies such as the western development campaign. In passing, it is noted that redistributive policies implemented in the past did not produce equalising effects as the role of transfer income is found to be inequality increasing (Wan et al. 2003).

Contrary to the declining share of geography, capital input contributes more and more to total inequality. Its contribution was negligible in the 1990s, typically around 2-4 percent. It increased to 16-24 percent in the new millennium.[7] In fact, the increase in total inequality in recent years can be largely accounted for by the increased contribution of the capital variable. This is in line with the modernization process of the rural economy in China. As the rural sector becomes more capital intensive and as capital becomes more unevenly distributed, its increasing share in total inequality is inevitable. Based on this finding, it is suggested that the government should give prime attention to credit services in rural areas, paying special attention to the poor. The provision of such services is important in terms of both income growth and inequality reduction. Interestingly, taking urban and rural China as a whole, capital input is also found to play a dominant and an increasing role in determining total inequality (Zhang and Zhang 2003).

[7] The jump in the contribution of the capital variable in 2000 might be partly caused by a launch of capital-intensive projects in some villages. While the magnitude of the jump is unexpected, this problem generated by data does not seem to affect our major findings.

Cropping pattern, represented by the variable 'Grain', is found to be a positive and important contributor to income inequality. Throughout the 1990s and up to 2002, this factor contributed almost 10 percent to total inequality. The percentage is larger than the contributions by labour input, human capital inputs or urbanization (as denoted by the variable 'Wage earner' in Tables 3 and 5). This finding implies that a pro-grain policy will help narrow income gaps. Policy initiatives to assist grain producers are likely to reverse the sign of this variable in the income generation function. As a result, grain may become an equalising factor rather than inequality-increasing factor. The reversal of the sign could mean significant reduction in the overall inequality. For example, if the sign of the contribution is reversed while magnitude maintained, the Gini value would come down by 0.1, which is rather substantial.

The next noticeable contributor is labour input. Its contribution is positive because per capita labour input implicitly captures the effects of the dependency variable. It is not difficult to infer that poor households have a higher dependency ratio (or lower per capita labour input) thus a lower income level. Inequality induced by this variable is likely to be transitory as an examination of rural population data indicates a converging trend in household size and dependency ratio. Thus, this positive contribution is expected to decline in the future.

Land is the only equalizing factor of inequality. This is understandable as land is known to be more abundant in less developed areas and those who are poorer are largely associated with farming. Unfortunately, the equalizing impact is negligible. To enhance this impact, policy initiatives are needed to increase returns to land and encourage land transfer to poor farmers. For a long time, economists have been arguing for the formation of land market in China, which could promote the land transfer. The very fact that land is collective-owned and cannot be traded constitutes a major obstacle to the establishment of a proper land market in China. Many households are reluctant to give up land because it acts as a security for livelihood in case of economic or political crisis. Therefore, in the near future, increasing returns to land would be more effective than enhancing land transfer as far as inequality reduction is concerned.

Substantial income gaps exist between households engaged in different business activities. The sector dummies are associated with a considerable share of total income inequality, signaling barrier to entry and constraints on resource movement between sectors of the rural economy. These may include institutional barriers (e.g., lack of transparent legal framework for granting business licenses) and economic barriers (e.g., accumulation of funds needed to set up companies). Adding education, age and training together, human capital contributes about 4-5 percent to total inequality. This small contribution implies either a small effect on income generation of human capital or less than expected uneven distribution of human capital across rural households in China. Nevertheless, the contribution is positive and human capital will play a more and more prominent role in rural economic growth as technology advances. On the other hand, economic reforms have eroded the state-funded education system and education gaps are enlarging between the rich and the poor at all levels in China. Therefore, the Chinese government must act quickly to address the access-to-education problem. Otherwise, it could become one of the major driving forces of income inequality in the not-too-distant future.

Referring to the second last row of each panel in Table 5, it is clear that our empirical model explains over 80 percent of total inequality as measured by the Gini index. The figure is smaller but still over 60 percent for other inequality indicators.

5 Conclusion

This paper combines the recently developed Shapley value framework of Shorrocks (1999) and the regression-based decomposition technique in analysing income inequality in rural China. The use of household-level data is complementary to most existing studies and the availability of time series data allows us to examine changes in total income inequality and its components over time. It is found that geography is the most significant contributor and will remain so in the future. Capital input has become a most important factor in affecting income inequality in rural China. The only equalizing variable is land input but its impact is minimal. Cropping pattern is more crucial than labour and human capital inputs in constituting total income inequality. It is suggested that China should endeavour to improve rural credit services and raise returns to grain-cropping in order to reduce inequality. The impact of education on inequality is small but is expected to grow. The current labour force enjoyed rather equal educational access prior to the reforms. As gaps in education have expanded in the last 15 years or so and as the rural economy becomes more skill demanding, the contribution of education to income growth and inequality is expected to increase.

References

Adams, R.H. (2002). 'Nonfarm Income, Inequality, and Land in Rural Egypt', *Economic Development and Cultural Change* 50:339-63.

Blinder, A.S. (1973). 'Wage Discrimination: Reduced Form and Structural Estimates', *Journal of Human Resources* 8:436-55.

Bourguignon, F. (1979). 'Decomposable Income Inequality Measures', *Econometrica* 47:901-20.

___, M. Fournier, and M. Gurgand (2001). 'Fast Development with a Stable Income Distribution: Taiwan, 1979-94', *Review of Income and Wealth* 47:139-63.

Chen, J., and B. Fleisher (1996). 'Regional Income Inequality and Economic Growth in China', *Journal of Comparative Economics* 22:141-64.

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.

Griffin, K., and A. Saith (1982). 'The Pattern of Income Inequality in Rural China', *Oxford Economic Papers* 34:172-206.

Griffin, K., and R. Zhao (1993). *The Distribution of Income in China*, St Martin's Press: New York.

Howes, S., and A. Hussain (1994). 'Regional Growth and Inequality in Rural China', *LSE Working Paper* EF/11, London School of Economics: London.

Hu, A., S. Wang, and X. Kang (1997). *Regional Disparities in China*, Liaoning People's Press: Shengyang.

Heltberg, R. (2003). 'Inequality in Vietnam: A Regression-based Decomposition', paper presented at the UNU-WIDER Conference on Inequality, Poverty and Human Well-Being, 30-31 May in Helsinki.

Juhn, C., K.M. Murphy, and B. Pierce (1993). 'Wage Inequality and the Rise in Returns To Skill', *Journal of Political Economy* 101:410-42.

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

___ and ___ (2005). 'Fifty Years of Regional Inequality in China: A Journey through Central Planning, Reform and Openness', *Review of Development Economics* (forthcoming).

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

Knight, J., and L. Song (1993). 'The Spatial Contribution to Income Inequality in Rural China', *Cambridge Journal of Economics* 17:195-213.

Lyons, T. (1991). 'Interprovincial Disparities in China: Output and Consumption, 1952-1987', *Economic Development and Cultural Change* 39:471-506.

Morduch, J., and T. Sicular (2002). 'Rethinking Inequality Decomposition, with Evidence from Rural China', *The Economic Journal* 112:93-106.

Oaxaca, R. (1973). 'Male–Female Wage Differences in Urban Labour Markets', *International Economic Review* 14:693-709.

Ravallion, M., and S. Chen (1999). 'When Economic Reform is Faster than Statistical Reform: Measuring and Explaining Income Inequality in Rural China', *Oxford Bulletin of Economics and Statistics* 61:33-56.

Rozelle, S. (1996). 'Stagnation Without Equity: Patterns of Growth and Inequality in China's Rural Economy', *The China Journal* 35:63-92.

Shorrocks, A.F. (1980). 'The Class of Additively Decomposable Inequality Measures', *Econometrica* 48:613-25.

____ (1982). 'Inequality Decomposition by Factor Components', *Econometrica* 50:193-211.

____ (1984). 'Inequality Decomposition by Population Subgroups', *Econometrica* 52:1369-85.

___ (1999). 'Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value', unpublished manuscript, Department of Economics: University of Essex.

____ and G. Wan (2004a forthcoming). 'Spatial Decomposition of Inequality', *Journal of Economic Geography*.

___ and ___ (2004b). 'A Method for Generating Income Data from Lorenz Coordinates', unpublished manuscript, UNU-WIDER: Helsinki.

SSB (State Statistical Bureau) (2003). *China Statistical Yearbook*, China Statistical Press: Beijing.

____ (various), *China Statistical Yearbook*, China Statistical Press: Beijing.

Tsui, K.-y. (1991). 'China's Regional Inequality, 1952-1985', *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.

____ (2002). 'Regression-Based Inequality Decomposition: Pitfalls and a Solution Procedure', *WIDER Discussion Paper* 101, UNU-WIDER: Helsinki.

____ (2004). 'Accounting for Income Inequality in Rural China: A Regression-based Approach', *Journal of Comparative Economics* 32:348-63.

____ and E. Cheng (2001), 'Effects of Land Fragmentation and Returns to Scale in the Chinese Farming Sector', *Applied Economics* 33:183-94.

____, Z. Huang and M. Wang (2003). 'Income Inequality in Rural and Urban China: The Role of Transfer Income', unpublished manuscript, UNU-WIDER: Helsinki.

Wei, H., K. Liu, M. Zhou, D. Yang and W. Hu (1997). *Regional Development in China: Economic Growth, Institutional Transition and Regional Inequality*, Business Management Publishing House: Beijing.

Wu, Y. (2000).'Income Disparity and Convergence in China's Regional Economies', in P.J. Lloyd and X. Zhang (eds), *China in the Global Economy*, Edward Elgar: Cheltenham.

Yao, S. (1997). 'Industrialization and Spatial Income Inequality in Rural China, 1986-92', *Economics of Transition* 5:97-112.

Zhang, X., and K. Zhang (2003). 'How does Globalization Affect Regional Inequality within a Developing Country? Evidence from China', *Journal of Development Studies* 39:47-67.