Hitotsubashi University

Global COE Hi-Stat Discussion Paper Series 019

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

The Contribution of Social Networks to Income Inequality in Rural China: A Regression-Based Decomposition and Cross-Regional Comparison

Ming Lu
Jianzhi Zhao

January 2009

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

Hi-Stat Discussion Paper

The Contribution of Social Networks to Income Inequality in Rural China: A Regression-Based Decomposition and Cross-Regional Comparison *

Ming Lu

Fudan University

and

Jianzhi Zhao

Fudan University

Abstract: This study aims to quantify the contribution of social networks, i.e., *guanxi*, to income inequality in rural households in China. One purpose is to understand how this influence varies across regions with different levels of marketization and economic development. Employing household survey data in rural China, we find that social networks contribute 12.1%–13.4% to income inequality among households in rural China, ranking fourth after village identifiers, nonfarm employment, and education. We also find that social networks exert a greater impact on income and a greater contribution to income inequality in Eastern China, compared with Middle–Western China where economic development is relatively slower. Our findings challenge the conventional understanding that social capital is the capital of the poor. In other words, the rich get richer in richer regions through social networks.

Keywords: Social Network, Income Inequality, Marketization, Regression-Based Decomposition

JEL Classification: O15; Z13; P36

---

* Ming Lu, Department of Economics, Employment and Social Security Research Center and China Center for Economics Studies, Fudan University, Shanghai, 200433, Tel: 86-21-55665308, E-mail: lm@fudan.edu.cn ; Jianzhi Zhao, Department of Economics, and Employment and Social Security Research Center, Fudan University, Tel: 86-21-55070563, Email: jianzhi.zhao@hotmail.com .We are grateful to the participants in the 2008 Annual Meeting of Chinese Economist Society and 2008 Hong Kong Economic Association Biennial Conference. We are indebted to Professors Zhao Chen, Ziyin Fan, Maarten Goos, and Bruce Meyer for their helpful comments and suggestions. All remaining errors are our own. Ming Lu acknowledges financial support from the National Social Science Foundation (08BJL008) and China Ministry of Education (07JZD0023). We also thank the School of Economics in Fudan University for "985" project and Shanghai Leading Academic Project (B101). Financial support for English proof editing from the Fukino Project, Hitotsubashi University is gratefully acknowledged. We also thank Prof. Hiroshi Sato for his help in preparing English manuscript.

1. Introduction

Social networks play an important role in the organization of societal life, especially in *guanxi* in China; i.e., social networks.[1] Empirical studies have found *guanxi* in Chinese society can bring about economic well-being through reductions in poverty (Zhang *et al.*, 2007) and higher income levels (Knight and Yueh, 2002), as well as facilitating the development of rural industry owing to the kinship network in rural China (Peng, 2004). This body of research, however, focuses on the benefits of an individual's social network without taking into consideration the influence of the distribution of social networks upon income inequality. Thus, it gives rise to the question whether inequality in social networks among households influences income inequality among households, and to what extent. Further, does this impact vary with economic development and marketization?

These questions are pertinent to the argument that "… social capital is the capital of the poor", as social capital, particularly social networks, facilitate credit access and contribute significantly to the welfare of the poor (Grootaert, 1999, 2001). Nevertheless, this argument could be somewhat ill-formed if the rich benefit more from their social capital, if the distribution of social capital favors the rich, or if the rich in richer regions receive relatively more than their peers in less-developed regions. All of these possibilities could undermine the argument that social capital is indeed the capital of the poor.

In an effort to clarify these questions, our paper studies the contribution of the inequality of social networks to income inequality in rural China using a newly developed method; namely, regression-based decomposition (Shorrocks, 1999). Our findings are as follows. First, social networks contribute some 12.1%–13.4% to total income inequality in our sample, following village identifiers, nonfarm employment, and education. This means that the rich benefit more

[1] In Chinese society, *guanxi* generally refers to one's own social network. This concept is widely used in academia; see, for instance, Li and Zhang (2003).

from social networks than the poor. Second, the contribution of social networks to income inequality in the East of China is greater than that in Middle and Western regions of China where the level of development and marketization lags behind. These findings imply that social networks, the foremost category of social capital, are more likely to be the capital of the rich. Moreover, the second finding may also question Stiglitz's (2000) argument that formal institutions, given the ongoing development of marketization, will increasingly displace the role of social capital as an informal institution. This cross-regional comparison likely predicts that social networks could play an even more significant role in the organization of societal life at more advanced stages of development. These findings then provide a better understanding of the nexus between social networks and marketization, as well as trends in and the transformation of the Chinese economy.

The structure of the remainder of the paper is as follows. Section 2 reviews the research literature with respect to social capital, income inequality, and the methods of analysis. Section 3 provides a description of the data and the variable selection. Section 4 undertakes an empirical study of household income determination and the regression results, followed by the regression-based decomposition of income inequality in Section 5. In Section 6, we group the data by geographical region and decompose the contribution of social networks to income inequality for each region. Section 7 provides some concluding remarks.

2. Literature Review

Since the late 1970s, with the implementation of the household responsibility system in rural areas of China, a multitude of income sources has arisen among villagers freed from collective agricultural production. It follows that the income inequality in rural areas has risen in the last 30 years. According to the World Bank,[2] the Gini coefficient in rural China increased by more than 60% from 1980 to 2005: from 0.25 in 1980 to 0.38 in 2005, almost at

[2] See the World Bank survey at http://iresearch.worldbank.org/PovcalNet/povcalSvy.html

the 'warning line' of inequality of 0.4. This increase in income inequality could jeopardize some aspects of economic development, including aggravating poverty (Wan and Zhang, 2006) and hamper income gains in rural households (Jalan and Ravallion, 2001). Hence, investigating the sources of the rising income inequality in rural China is not only significant but also imperative.

Research into income inequality in rural China has been the subject of increasing attention in academia and the source of many important findings. Existing research mainly focuses on the sources and determination of income, including the role of physical, human, and political capital. For example, Xu *et al.* (2008) show that the disaggregation of collective lands in the late 1970s led to land inequalities, which in turn was responsible for the rising income inequality among rural households. Likewise, Wang (2006) suggests that education plays a significant role in income inequality in rural China. Meanwhile, there are an increasing number of studies concerning the role of political capital, mainly represented by party members and village/town cadres. Analyzing data on rural households in Shandong Province, Morduch and Sicular (2000, 2002) concluded that being a party member or cadre exerts a positive impact on income (Morduch and Sicular, 2000) and contributes to a certain level of income inequality (Morduch and Sicular, 2002). Walder (2002) reveals similar results. Nevertheless, there has been lack of concern about the contribution of social networks to income inequality, especially on the contribution of the kinship networks that are an indispensable part of Chinese daily life.

The role of social capital, as initially proposed by Jane Jacobos (1961),[3] has increasingly drawn the attention of economists, sociologists, and political scientists. Although there is still some controversy about the exact definition of social capital, the seminal concept proposed by Putnam (1993) is that social capital refers to the collective value of all 'social networks' and

[3] Jacobos initially proposed the concept of social capital in her seminal text *The Death and Life of Great American Cities,* where she referred to the networks among the New Yorkers as an indispensable part of the flourishing of the city, although she did not explicitly define social network or social capital.

the inclinations arising from these networks to ‘do things for each other’. Most recently, there is an emerging consensus about the definition of social capital advanced by the World Bank (2000) that social capital refers to “… norms and networks that enable collective actions”. One can think of social capital, particularly social networks, as an asset people boast about and invest in to assist each other. In turn, studies have linked social networks to improvements in income (Narayan and Pritchett, 1997), the facilitation of employment (Montgomery, 1991; Munshi, 2003, 2006), especially for the alleviation of poverty (Grootaert, 1999, 2001) and better access to credit through the role of informal insurance (Grootaert, 1999; Bastelaer, 2000).

Using quantile regression, Grootaert (1999, 2001) draws the conclusion that ‘social capital is the capital of the poor’ on the basis that the return to social capital for people at the 10^{th} percentile of income is higher than that at the 90^{th} percentile. This proposition, however, is perhaps rather too hasty in aggrandizing the positive effect of social capital for the poor, mainly because Grootaert focused on comparing the return to social capital between the top and the bottom percentiles without taking into consideration the distribution of networks between these groups, and did not consider income decomposition for further study. In Grootaert’s paper, the rich enjoy relatively more social capital: this suggests that social capital is more likely the capital of the rich. In an effort to address these puzzles, our paper focuses on social networks, the most vital measurable category of social capital at an individual level, to see whether social capital is indeed the capital of the poor. Moreover, the data used in the existing literature either is collected in developed countries (like the US) or in countries at very low levels of development (e.g., rural areas in Tanzania). However, China, a country experiencing breathtaking development and transformation with great disparity across different areas, can serve as an ideal research object for comparative study of the contributing factors to income inequality. This is because there is no need to be concerned about the availability and comparability of data in, say, an international comparative study.

The significance of *guanxi* or social networks is appreciable in the daily life of the Chinese people and thus attracts considerable interest from academia. *Guanxi* has been seen to promote employment, not only for migrants (Li and Zhang, 2003), but also urban residents (Bian, 2001). Moreover, there is already some work concerning the influence of *guanxi* upon income inequality between groups. For instance, Lin (2001) considers the influence of social network inequality from the perspective of capital. Lin (2001) argues there are two channels through which social network inequality affects income inequality: capital deficits and return deficits. Capital deficits are the extent to which different social groups, for reasons of investment or opportunities, possess a different quality or quantity of capital, while return deficits are the extent to which a given quality or quantity of capital generates differential returns for different social groups. Using labor market data for urban China, Lin (2001) finds one reason for the higher income of male workers is not only that they have greater access to social capital than female workers, but even for the same quantity of social capital the return to social capital is higher among male workers. The most severe drawback of Lin's (2001) method is that it is limited to a certain extent by the grouping of the sample. Lin (2001) also failed to check other sources or factors that may contribute to income inequality.

In contrast with the static analysis of social networks, revealing how they influence social network changes in the process of development and marketization is a particularly exciting and challenging aspect of development economics. According to Stiglitz (2000), formal institutions will displace social capital, as an informal institution, during the development process. Several scholars have thus given attention to the role of traditional social networks in the process of development, transformation, and globalization. Munshi and Rosenzweig (2006), using survey data on school enrolment and income in Mumbai, found that the traditional caste-based social network channelled lower-caste boys into local language schools followed by traditional occupations, whereas lower-caste girls, who did not benefit from caste-based social networks, could exploit opportunities provided by globalization to switch to

English language schools.

Through study of the labor market in urban China, Knight and Yueh (2002) and Li, Lu and Sato (2008) find that social networks not only can exert a positive influence on income, but that this influence is larger in private sector firms than in state-owned enterprises. Based on the premise that the private sector represents the trend in marketization and development, they draw the conclusion that social networks become even more significant with the transformation of the economy. Conversely, some studies have found that the social network function decreases with the development of formal institutions. Zhang *et al.* (2007) suggest that household level social networks have less effect on poverty reduction in areas with a high level of marketization. In addition, Lu *et al.* (2008) conclude that with the growing opportunities to work outside villages, social capital is losing its ability to cushion natural shocks for rural households in China. Following this literature, our paper attempts to discover how the influence of social networks on income inequality is evolving in response to the processes of marketization and economic development.

In the empirical analysis of income distribution, there are three main approaches: semiparametric and nonparametric methods, decomposition by grouping, and decomposition by regression. The semiparametric or nonparametric techniques, first proposed by Deaton (1997) and Dinardo *et al.* (1996), attempt to analyze income inequality through the distribution of the income density function. These approaches, however, are less than persuasive because they lack restrictions on the model of income determination. The second method is to group data in accordance with the sample characteristics, such as education, gender, etc. As pointed out by Morduch and Sicular (2001), the drawbacks of this method lie in three key areas. First, this method cannot apply to continuous variables (such as age) that are often quite important and necessary in the analysis of income inequality. Second, there are limitations in the number of variables that can be included because the number of groups will multiply with the number of categories for each factor. Finally, it is impossible to control for the endogeneity of the variable used to group data, presumably because the researcher groups data with an exogenous variable.

Accordingly, the most popular method used in researching income inequality is the decomposition of inequality with regression. Oaxaca (1973) and Blinder (1973) originally proposed this line of research to decompose between-group differences in the mean of income. In more recent times, Field and Yoo (2000) and Morduch and Sicular (2003) have greatly developed this line of analysis by developing their own methods of decomposition. However, there are several limitations in their work. First, they confine their regression models to only a few specific forms: Field and Yoo (2000) use a semilog model, while Morduch and Sicular (2003) assume a linear model. Second, these studies are also restricted in adopting a specific measure of income inequality. Finally, the contribution of the residual term to income inequality remains unexplained. However, the regression-based Shapley value decomposition proposed by Shorrocks (1999) addresses these limitations, later refined by Wan (2004), to tackle the contribution of the residual term.[4]

In sum, the main contributions of our paper are twofold. First, we quantify the contribution of social networks to income inequality using the regression-based decomposition method. Second, by utilizing the stunning disparity between the Eastern and Middle–Western regions in rural China, our paper examines how the contribution of social network to income inequality varies with different stages of economic development.

3. Description of the Data and Variables

The main data source in our paper is the *China Rural Survey 2004* conducted by the China Center for Economic Studies at Fudan University. The sample includes information on 927 households in 48 villages across 22 provinces in 2003, comprising approximately 20

[4] For a review of income inequality analysis methods, especially income inequality decomposition, see Morduch and Sicular (2002) and Wan (2004).

households in each village and two villages in each province. The most striking advantage of this dataset is its rich information about social capital in rural China, particularly social networks.

We group the data according to the level of marketization from the *Report on China Marketization Index 2002* (Fan and Wang, 2004). This report builds an evaluation system for the relative level of marketization of provinces in China using several characteristics, including the relationship between government and the market, the development of the private sector, the development of the product and factor markets, and the development of formal institutions and the legal system. The relative marketization indices grade continuously from one to 10 so that the higher the index, the deeper the process of marketization. We use the marketization index from 2002 to avoid any potential simultaneity with the other variables. Figure 1 depicts the marketization level by province in China. It is clear that the development of markets in different regions is unbalanced, as shown by the variance in the level of marketization. The numbers in parenthesis for each province show (from left to right) the respective number of villages, the number of households, and the level of marketization.

[Figure 1: Here]

Table 1 lists the variables and their definitions. The dependent variable is the natural logarithm of household income per capita from summing “income from land, forest, stock, fruit, and fishing”, “income from nonagricultural occupation in local areas”, “income from employment outside county”, and “income from property”. The independent variables are categorized into family characteristics, household physical capital, human capital, political capital and village dummy variables, all of which are generally found in existing research (e.g., Morduch and Sicular, 1999, 2002; Walder, 2002; Wan, 2004; Wan *et al.*, 2006). More importantly, as social networks are the focus of our paper, we explain its measurement as follows.

[Table 1: Here]

Because the most vital social network formations in China are the kinship network and the close friendship relationship (Knight and Yueh, 2002), we mainly focus on two social network measures. First, two questions in our survey are "How many relatives and close friends in the city do your family have?" and "How many relatives and friends do your family have in the local government area?" We add the responses to these two questions to measure the social network of a family. Second, we employ the ratio of "the amount you spend in purchasing gifts for your relatives or close friends at Chinese New Year" and "the money you spend on weddings, birthdays and funerals of your relatives or close friends"[5] to obtain the total expenditure on daily life.[6] We exclude occasional expenditures, such as expenditure on durable commodities or housing construction, to avoid measurement error.

The rationale for the second measure is from Yan (1996) who finds that rural households in northern China use more than 20% of their total expenditure in exchanging gifts to maintain *guanxi.* However, this may exaggerate the cost of social networks because it does not exclude situations when people invest in *guanxi* only for short-term interests, such as finding a job, which could sharply increase the expenditure on gifts. Hence, it is more reasonable and justifiable to take into consideration only holiday or ceremony expenditure as this expenditure is more smooth and continuous. Further, the ratio of gift expenditure to daily expenditure in place of its absolute value can help mitigate the simultaneity problem in *guanxi* as the rich will have higher daily expenditure so they may spend more on gifts, and vice versa. To check for any potential problems with multicollinearity between the social network variables, we

[5] According to Chinese tradition, people should send gifts to relatives and friends on Chinese New Year, as well as on other occasions, such as weddings, birthdays, and funerals.
[6] For daily expenditure, we add the expenditures on food, cigarettes and alcohol, transportation fees, phone fees, utility fees, and other daily expenditure.

checked the covariances between them. The correlations are 0.11, 0.17, and 0.09 for the total, Eastern, and Middle–Western samples, respectively. This removes the possibility of severe multicollinearity in the regression.

In Table 2, we provide a description of the data; first the total sample, and then after categorizing them into Eastern and Middle–Western regions. It is evident that the values of some variables vary across these regions. In terms of the dependent variable, the natural log of household income per capita, the mean is higher in the East than in the Middle–West: this is consistent with our intuition. However, the standard deviation is the reverse, meaning that income inequality is higher in the East than in the Middle–West. For the two *guanxi* variables, the number of "relatives and close friends in the city and local government area" in Eastern households is no higher than in Middle–Western households, and the lower ratio of gift expenditure to daily expenditure in the East indicates we have mitigated the endogeneity of the *guanxi* variables. In an effort to guarantee the exogeneity of the variable for land, we exclude rented land.[7] The land per capita in the East is a little higher than in the Middle–West because some villages in eastern provinces, such as Hebei, have a larger land endowment. As for the household characteristics, the number of people in the Eastern sample is much less than that in the Middle–Western sample. This is perhaps the result of tighter enforcement of the "one child policy". In addition, the nonagricultural employment rate in the East is much higher than that in the Middle–West: this is because the East is generally more industrialized. More interestingly, the average years of education for Eastern workers is greater than for their peers in the Middle–Western areas. However, the converse holds for standard deviation, which could be the result of a higher dropout rate or an earlier employment age for students in Middle–Western areas.

[7] Under the household responsibility system, land is distributed to households based on the number of heads. Farmers do not own their land but they can rent it from or to others. However, given the possibility that households can rent more land from or to others, we simply include the lands allocated by village. This is completely exogenous with respect to income.

[Table 2: Here]

4. Empirical Model and Regression Results

The regression-based decomposition developed by Shorrocks (1999) comprises two steps: (i) setting up a regression model and estimating the coefficients; and (ii) decomposing the income inequality indices based on the regression estimates. First, we estimate the empirical model as follows:

$$Lny_{ij} = a_0 + \beta_1 SN_{ij} + \beta_2 FC_{ij} + \beta_3 OC_{ij} + \beta_4 HC_{ij} + \beta_5 PC_{ij} + \beta_n FE_j + u$$

where LnY_{ij} denotes the natural log of the income per capita of household i in village j, SN is the measure of social network, FC is a set of family characteristic variables, OC denotes the set of physical variables—mainly land per capita, HC is the abbreviation of human capital—including the education and age of workers, and PC is household political capital (measured by membership of the Communist Party). Meanwhile, given the myriad uncontrollable village characteristics that may influence household income, it is reasonable to control for these using village dummies. We use a semilog regression as this can render the dependent variable closer to a normal distribution, as is common in income determination models.

Table 3 reports the results of the estimation. In Model I, we exclude the *guanxi* variables. In Model II, the network measurements are included. After comparing the results of these models, we note that the variables that are significant in Model I are also significant in Model II and with the exception of party membership, their coefficients do not change substantially, though the R^2 increases by 0.03.

[Table 3: Here]

With respect to the social network variables, the results show the estimated coefficients are all significantly positive, indicating they are positively correlated with the level of household income. Moreover, we find that if a household has one more relative or close friend in the city or local government area, then household income per capita increases by 6%.[8] In addition, if the ratio of expenditure on gifts to daily expenditure grows by 10%, there will be a 4.5% increase in household income per capita.[9] These results illustrate that the influence of *guanxi* on income determination is considerable, although omitting these variables does not substantially bias the estimation of the remaining coefficients. However, with income decomposition analysis, omitted variables will exaggerate the contribution of the other independent variables and the error term.

In terms of family characteristics, workers per capita play a positive role in income determination. The ratio of male workers is statistically significant; that is, in rural households, male income per worker is higher than female income per worker. The coefficient for the nonagricultural employment ratio is not only significant but also large: a 10% increase leads to an 8.8% growth in household income per capita. Per capita land has a significantly positive coefficient, consistent with previous research in this area (e.g., Morduch and Sicular, 2000, 2002; Wan and Zhou, 2005). As for the human capital variables, while the coefficient of workers' average age is statistically insignificant, the average education of workers exerts a significant inverse U-shaped influence on the level of income. With the political capital, the coefficient for the average party member is significantly positive in Model I but insignificant in Model II, perhaps because there is a correlation between political capital and the number of relatives or close friends in the city or local government area (the covariance between these

---

[8] We calculate the partial influence using $\%\hat{y} = 100(e^{b} - 1)$. For details, see Wooldridge (2003).

[9] Given that the return to the ratio of gift expenditure could be an inverted U-shape, we check this by adding its square to the regression model. The results show that the square term is not significant and therefore we exclude it from the final regression model.

variables is 0.18). Hence, the omitted variable bias in Model I indicates that the party member variable may actually exercise no significant influence on income as it partly represents other unobservable factors that could also affect household income. This argument, to some extent, is consistent with Li *et al.* (2007) who study samples of twins and find party membership has no significant effect on income. Moreover, the covariance between party membership and the ratio of gift expenditure to total daily expenditure is only 0.09, suggesting a low probability of increasing gift expenditure for party members.

5. Regression-Based Decomposition of Income Inequality

In the shaping of income inequality, there are two main channels deciding the contribution of a certain factor to the inequality index. First, the coefficient of this variable could have a positive influence on the income inequality given its distribution. The greater this influence, the more unequal the income distribution. Second, if the returns to this variable are the same, its distribution decides the total income inequality index. That is, the more uneven the distribution of the variable, the greater its contribution to income inequality. If the distribution of the factor is equally distributed or its coefficient is zero, then its contribution to income inequality is zero. This is the core conceptualization of regression-based decomposition.[10]

Our paper employs the Shapley value decomposition framework proposed by Shorrocks (1999) to calculate the contribution of each variable to total income inequality. The method involves intensive computing requirements.[11] Suppose $Y = f(X_1, \ldots, X_K)$ is a general income-generation function. Usually X is different for different individuals. Replacing the real value of X_k with its sample mean would eliminate any differences in X_k among individuals. It is easy to recompute Y after this replacement. The resulting income, denoted by Y_k, differs from

---

[10] Regression-based decomposition first appeared in Oaxaca (1973), in which he argues that gender income inequality is due not only to different salaries for the same position for male and female workers, but that the distribution of working opportunities is uneven for males and females.
[11] The World Institute for Development Economics Research of the United Nations University (UNU-WIDER) developed a Java program in light of this problem.

individual to individual because X, other than X_k, differ with different individuals. However, the differences can no longer be attributed to X_k; i.e., inequality in Y_k, denoted by $I(Y_k)$, is brought about by variances in X except X_k. According to the most natural rule in Shorrocks (1999), the contribution of X_k to total inequality, C_k, can be obtained as $I(Y) - I(Y_k)$ for $k = 1, \ldots, K$. 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 $k, j = 1, \ldots, K$ ($k \neq j$). Similarly, the third-round contribution can be obtained as $C_k = I(Y_{ij}) - I(Y_{ijk})$ for $k, j, i = 1, \ldots, K$ ($k \neq j \neq i$). This process continues until all X are replaced by their sample means. In each round, it is possible to have multiple C_k that are first averaged and then averaged across all rounds.[12]

Because we employ a semilog model in the income determination procedure, in an attempt to avoid the distortion of inequality index, it is necessary to transform it back to a linear model to acquire the real value of Y, as follows:

$$y_{ij} = \exp(\hat{a}_0) \bullet \exp(\beta_1 SN_{ij} + \beta_2 FC_{ij} + \beta_3 OC_{ij} + \beta_4 HC_{ij} + \beta_5 PC_{ij} + \beta_n FE_j) \bullet \exp(\hat{u})$$

where $\exp(\hat{a}_0)$ denotes the constant term, which is ruled out in the decomposition of income inequality because it is equal for all samples. As for the residual term, because we are unable to analyze its distribution, it is prudent for us to examine its contribution to total income inequality. Following Wan (2004), we obtain this contribution by comparing the original income inequality with the predicted value. The ratio of the former to the latter is the extent to which the variables in our regression model can explain total income inequality. Strictly

[12] A further explanation of the Shapley value is offered in Appendix I. For more details about the approach, refer to Shorrocks (1999).

speaking, if the contribution of the residual term to total inequality is nil (i.e., the variables in the regression model can entirely interpret the inequality), the ratio will be 100%. Moreover, in an attempt to check the robustness of our results, we employ three commonly used income inequality indices: the Gini coefficient, the Atkinson, and the generalized entropy (GE) indexes.[13] Table 4 tabulates the calculated results of the explained contribution of the variables in the regression model in the form of a percentage for all three indices and all samples. It is evident that all of the income inequality measures used explain more than 50% of the variation in household income, especially the Gini coefficient, which explains more than 80%.

[Table 4: Here]

As previously discussed, the calculation of the Shapley value decomposition involves an enormous volume of computing power, which makes it difficult to calculate the decomposition process for more than about 10 explanatory variables. To simplify the computation, we combine variables belonging to the same category: (1) all village dummies with their coefficients sum to a single factor; (2) the number of family members and workers per capita comprise one factor; and (3) a single variable is used to represent education and its square in the decomposition procedure. In doing so, we can acquire the decomposition results with little concern for the absence of the contribution of the principal variables to income inequality.

Table 5 tabulates the decomposition results based on the regression model. The left column is the percentage contribution of a given variable to total income inequality, while the

[13] The function of these three indices are as follows: $Gini = \frac{2}{n^2 u}\sum_{i=1}^{n}(i = \frac{n+1}{2})y_i$, $Atkinson = 1 - \prod_i (\frac{y_i}{u})^{1/N}$, $GE = \frac{1}{a(1-a)}\sum_{i=1}^{n} f_j[1-(\frac{Zj}{\mu})^a]$, where y is income, u is the mean of y, i denotes the sample (i =1, 2,…, N) and f_j is the proportion of samples. In the GE, a is a constant indicating the degree of inequality aversion. The larger this number, the greater the degree of inequality aversion. In our analysis, we define a = 0 and then acquire $GE_0 = \frac{1}{n}\sum_{i=1}^{n}\ln(u/y_i)$, the so-called Theil–L index.

right column is the ranking of this variable among all variables. Additionally, we note that the contribution of a given variable may vary across the different income inequality indices. This is because different indices are associated with different social welfare functions that assume different aversions to inequality. In addition, they place different weights on different segments of the underlying Lorenz curve. Therefore, different indices of inequality often produce different measurement results, and these may carry over to the inequality decomposition. Nevertheless, the top five variables are identical for each index and their ranking is unchanged, suggesting that the contribution of these variables to income inequality is rather robust. Though the ranking of the other five variables are inconsistent across the inequality indices, they are the same for the last two indices. Allowing that the contribution of the final four variables is considerably smaller than the first five factors, it is no exaggeration to draw the conclusion that our decomposition is rather robust and convincing.

[Table 5: Here]

To begin, the results shows the aggregate contribution of the two network variables are more than 12%, even up to 13.4% for the Gini coefficient, suggesting inequality in *guanxi* actually exerts a significant effect in the shaping of income inequality. More specifically, the number of relatives and close friends in the city or local government area contribute more than 9% across the three inequality indices, higher than any other variable ranking below fifth. Additionally, another *guanxi* variable—the ratio of gift expenditure to total daily expenditure—contributes more than 2% to the Atkinson index and GE_0, and up to 3.8% to the Gini coefficient.

In light of the great differences in the contribution to income inequality across the *guanxi* variables, it is necessary and even interesting to analyze the possible reasons. We have already discussed that there are mainly two reasons determining the contribution of a certain factor to income inequality: the coefficient and the distribution of the variable. That is, the bigger the coefficient, the greater the contribution and the more uneven the distribution, the more significant the contribution. Hence, we deduce that the main reason for the great contribution of the number of relatives or close friends in the city or local government area lies in its highly uneven distribution, because the estimated coefficient is only 0.057 while its standard deviation is as high as 3.08 (bearing in mind the mean value is only 2.21). On the other hand, the other *guanxi* variable—the ratio of gift expenditure to total daily expenditure—contributes to inequality largely because of its great influence over income determination, whereas its distribution is relatively equal when compared with its mean value. These two contrasting scenarios are interesting when allowing for the following possibilities. The distribution of the former variable is quite exogenous for rural households. For the ratio of gift expenditure, however, households can invest more on their *guanxi* as long as they realize the profits of their investment. Thus, its relatively smaller contribution is most probably due to the equal distribution of this variable, even though the coefficient is rather large.

The village dummies rank first across all factors contributing to income inequality in all indices, consistent with our intuition that there are myriad of unobservable disparities between villages in China that will unquestionably influence household income, including location, natural endowment and infrastructure. This finding is consistent with the previous literature. The contribution of the space factor in overall income inequality in China has been emphasized by most previous studies, including Wan (2004), Gustafsson et al. (2008), and Yue et al. (2008). Moreover, some recent studies on Chinese regional income disparities pay attention to economic variance at the lower regional level (counties, townships, and villages). For example, Morduch and Sicular (2002) illustrated the importance of the location factor among different counties in the same region. Using a village survey conducted in the Handan district, Knight and Li (1997) found a 'cumulative causation' that resulted in economic disparities among villages in the same district. Likewise, Perkins (2003) employed filed research and data collection in suburban Tianjin to demonstrate that large economic variations existed among villages in the same township, including substantial differences in economic structure and the level of well-being. Sato (2003), based on field research, proposed a typology of villages and examined the relationship between village characteristics and peasant income. Using data on 961 villages distributed across 22 provinces, Sato (2008) also demonstrated that village-specific socioeconomic factors, including social, physical and human capital, strongly influenced household income.

As for the nonagricultural employment rate, it ranks second in all inequality indices and its contribution is more than 20%. This is quite appreciable given that the returns in nonagricultural occupations are higher than in agricultural work, especially in those areas with barren land or a lack of natural endowments. The contribution of the education factor is surprisingly high, ranking third across all indices with a higher than 18% contribution. This partly accords with Morduch and Sicular (2002), where they find that education contributes more than 16.9% to the Gini coefficient of total income inequality in rural China. For this reason, our results once again prove the significance of education in rural China and provide evidence for the justification of compulsory education to narrow income inequality.

The ratio of male workers in the work force is also quite important, ranking fifth across all indices with a contribution greater than 5%, suggesting that there is still great income unevenness between male and female workers in rural households. For the combined variable of the number of people and workers per capita, they are only positive in the Gini coefficient but negative elsewhere, and this probably arises from the convergence in family size and workers per capita brought about by the strict “one child policy”. Wan (2004) also finds that the disparity in the number of people and worker ratio is narrowing in rural China.

The contribution of land per capita is rather small across all three indices, less than 1%, and even negative in the Atkinson index, suggesting land only partially narrows income inequality in rural Chinese households. This is also consistent with the high contribution of the nonagricultural employment ratio to income inequality, as families with less land would choose to work outside in nonagricultural occupations that will actually compensate for the loss of land, and thus bridge the gap between themselves and people with more land. Additionally, with the adoption of the household responsibility system in rural China, all lands are allocated per capita, and this can also explain the low contribution of inequality in land to income inequality.

The role of party membership is rather trivial in our results with a contribution of less than 0.6% to each index. This finding is consistent with Morduch and Sicular (2002), who find that the number of party members has little influence on income inequality (for example, only 0.42% to the Gini coefficient). The average age of the workforce, used to estimate the sophistication of workers, is also weak. This suggests that while the increase in age would probably imply greater experience, it also decreases the physical condition for rural households.

6. A Cross-Regional Comparison of the Role of *Guanxi*

In this section, we test how the contribution of *guanxi* to income inequality varies across areas at different stages of development by grouping the data into Eastern and Middle–Western areas by the level of marketization. We define Eastern areas as those with a level of marketizaton above 6.0, including Hebei, Liaoning, Beijing, Shandong, Jiangsu, Zhejiang, Fujian, Guangdong, and Hainan, all of which have benefited from their location and policy priorities in the process of reform and opening-up. The remaining data are categorized as Middle–Western, including Henan, Hubei, Guangxi, Gansu, Inner Mongolia, Shanxi, Jiangxi, Shaanxi, Hunan, Jilin, Ningxia, and Heilongjiang. Among these, all marketization indices are less than 6.0, except for 6.03 in Sichuan. However, allowing for the fact that Sichuan has a large ratio of agricultural workers and geographically belongs to the Western area, we still categorize it as a Middle–Western area.

(1) The income determination model and regression result

The income determination model is identical with the former one and we acquire the regression result in Table 6.[14]

[Table 6: Here]

The coefficients for both *guanxi* variables are higher in the East than the Middle–West. An increase in one relative or close friend in the city or local government area could bring about as much as 6.5% growth in income per capita in Eastern areas, but only 5.3% in Middle–Western areas. Likewise, a 10% increase in the ratio of gift expenditure to total daily expenditure causes per capita income to rise by 8.9% for households in the East and only 3.1% for their peers in the Middle–West. These findings are consistent with Lin's (2001) argument for a 'return deficit': that is, one source of social capital inequality is that a given quality or quantity of social capital brings about differential returns for different individuals or groups.

(2) The inequality decomposition results

The disparity in returns to *guanxi* across different areas is not enough to judge how its contribution to income inequality changes at different stages of development. Hence, we decompose income inequality using the regression model and rank the contribution of all factors for both areas, as tabulated in Table 7.

[Table 7: Here]

As shown, though the contribution of each factor varies across the inequality indices, the

[14] To test the necessity for grouping, we conduct an F-test of the Eastern and Middle–Western areas; the results show there is a systematic difference between these groups. The appendix includes details.

top five factors are identical and their rankings remain unchanged. In addition, despite the fact that the rankings of the final four variables change across the three indices, it is safe to say the decomposition results are robust and persuasive as the contribution of the final four factors is quite small when compared with the top five factors. However, the total contribution of the *guanxi* variables remains fourth in each index for both categories of data; their contribution in Eastern areas amounts to 17.8%, substantially more than the 12.9% in the Middle–Western areas. Likewise, for the Atkinson index and GE_0, the gap in the total contribution of *guanxi* between the East and Middle–West is large, some 3.9% and 5.1%, respectively.

The contribution of relatives or friends in a city or local government area in the East is a little smaller than in the Middle–West. This is mostly because the distribution of this factor is much more uneven in Middle–Western households: the standard deviation in Middle–Western households is 3.34 and only 2.67 in Eastern households. While the coefficient is larger for Eastern households, this does not mitigate the effects of the uneven distribution of relatives or friends in Middle–Western areas. Nevertheless, this does not mean the role of relatives or friends in shaping income inequality in more developed areas is less significant than in less developed areas. In fact, it is the reverse. We should not neglect the possibility that with further economic development and marketization, the contribution of relatives or close friends in a city or local government area could be higher with a larger coefficient; i.e., more considerable benefits.

The contribution of the *guanxi* variable—the ratio of gift expenditure to total daily expenditure—for Eastern households is as high as 8.1%, much greater than the contribution of only 2.7% in Middle–Western families. This is most likely because the coefficient for this variable is much greater in the East, as the distribution of this factor is relatively equal with a low standard deviation. We conjecture the main reason for this result lies in that *guanxi* could bring about benefits in the form of sharing information or mutual help in Eastern areas where there are greater economic opportunities.

The village dummies unsurprisingly remain the factor that contributes most to income inequality. We know household income inequality can be divided into two parts: inequality between villages and inequality within villages. The contribution of the village dummies indicating the inequality between villages is much higher in Middle–Western areas, suggesting the disparities between villages are much smaller in the East. The contribution of nonagricultural employment is more significant for Middle–Western households, up to 23.7% higher than the 18.4% for Eastern households. The main reason for this, as shown in Table 2, is that the coefficient for Middle–Western households is much higher as the standard deviation is not very different between the two areas.

The contribution of education to income inequality is not only much higher in the East, but rises faster in its ranking, especially for the Atkinson index and GE_0, where it jumps to the number-one factor. Combined with Table 2 and Table 6, we can deduce that it is because the coefficient of education for Eastern workers is much higher than for their counterparts in the Middle–West, as the standard deviation is much lower in the East than in the Middle–West. This suggests the consequences of education are more striking with economic development, even for the rural households who are traditionally a group with lower returns to education.

An interesting finding is the contribution of the ratio of party members, which is negative in Middle–Western areas but positive in Eastern areas. We argue that party membership stands for political capital, which can only bring about benefits when there are economic resources in the village. Therefore, for villages in Middle–Western areas, it is not only that there are fewer economic resources, but also because party members should spend their personal time on village affairs or not working outside the village, both of which could handicap income growth. In Eastern villages, however, there are more economic resources or even rents that will make it easier for party members to acquire access to income, leading to considerable returns to their political capital.

The ratio of male workers contributes more in Middle–Western areas than in Eastern areas because it is more unevenly distributed in the former, as indicated in Table 2, though the return is lower.

7. Conclusion

Our paper quantifies the contribution of *guanxi* or social networks to income inequality in rural China and shows how this contribution varies among areas at different stages of development and marketization. Our main findings are as follows.

(1) For rural households, *Guanxi* can significantly improve their household income, and inequality in its distribution will contribute to total income inequality as much as 12.1%–13.4%, ranking fourth behind village dummy variables, nonagricultural employment, and education. Hence, it is reasonable for us to defend social capital as perhaps not simply 'the capital of the poor', since at least social networks, the foremost category of social capital, widens the income gap between the poor and rich.

(2) *Guanxi*, as an informal institution, receives more benefits and is much more important in the shaping of income inequality for households in areas with greater development and level of marketization.

Because the data we employ in this paper are collected from rural China, we are not sure whether the same conclusion could still be drawn in Chinese cities. Moreover, *guanxi* in our paper mostly refers to 'strong ties',[15] so whether the results still hold for 'weak ties' remains unknown. Whether or not weak ties that are important in developed countries are also significant in contemporary Chinese society deserves further study. The more tantalizing question is whether the fact that *guanxi* are more important to households in Eastern areas will change if formal institutions develop, and if they will become increasingly mature in the future. These questions represent the scope of our future study.

---

[15] According to Lin's (2001) definition, we should categorize relatives or strong close friendships as strong ties.

Comprehending the evolution of the social network function could provide a better understanding of the trends and formation of future reforms. It will determine whether China will build a fully-fledged free-market economic system or a market embedded with nonmarket forces.[16] The undeniable reality is, however, that family-based *guanxi* is unevenly distributed, which could engender that 'not all men are created equal'. If there is nothing we can do about this inherited inequality, we should take measures to avoid it evolving into a further disadvantage for people later in their lives. After all, most of us are more willing to live in a society with equal opportunities, transparent regulations, and fair competition. Thus, how nonmarket forces evolve in the ongoing transformation and economic development of China is a topic deserving of further study.

[16] Stiglitz (2000) holds the view that the successful development of the Chinese economy in the past is partly owing to well-preserved and well-developed social capital.

References

[1] Bastelaer, van Thierry. 2000. *Does Social Capital Facilitate the Poor's Access to Credit?—A Review of the Microeconomic Literature*. Social Capital Initiative No. 8. Washington DC: World Bank.

[2] Bian, Yanjie. 2001. "Guanxi Capital and Social Eating in Chinese Cities: Theoretical Models and Empirical Analyses", in Nan Lin, Karen Cook, Ronald S Burt (eds.), *Social Capital: Theory and Research*, New York: Aldine de Gruyter.

[3] Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates". *Journal of Human Resources*, 8(4): 436–455.

[4] Deaton, Angus. 1997. *The Analysis of Household Surveys*. Baltimore: John Hopkins University Press.

[5] Dinardo, John, Nicole M. Fortin and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach", *Econometrica*, 64(5): 1001–1044.

[6] Grootaert, Christiaan. 1997. "Social Capital: the Missing Link?", in World Bank, *Expanding the Measure of Wealth-Indicators of Environmental Sustainable Development*. Washington DC: World Bank.

[7] Grootaert, Christiaan. 1999. "Social Capital, Household Welfare and Poverty in Indonesia". *Local Level Institutions Working Paper* No. 6. Washington DC: World Bank.

[8] Grootaert, Christiaan. 2001. "Does Social Capital Help the Poor: a Synthesis Findings from the Local Level Institutions Studies in Bolivia, Burkina Faso and Indonesia". *Local Level Institutions Working Paper*, No. 10. Washington DC: World Bank.

[9] Gustafsson, Björn, Li Shi, Terry Sicular, and Yue Ximing. 2008. "Income Inequality and Spatial Differences in China, 1988, 1995, and 2002", in Björn Gustafsson, Li Shi, and Terry Sicular (eds.), *Inequality and Public Policy in China*, New York and Cambridge: Cambridge University Press, 35–60.

[10] Fan, Gang and Xiaolu Wang. 2004. *Report on China Marketization Index in 2002* (in Chinese). Beijing: Economic Science Press.

[11] Fields, Gary S and Gyeongjoon Yoo. 2000. "Falling Labor Income Inequality in Korea's Economic Growth: Patterns and Underlying Causes". *Review of Income and Wealth*, 46(2): 139–159.

[12] Jacobos, Jane. 1961. *The Death and Life of Great American Cities*. New York: Random House.

[13] Jalan, Jyotsna and Martin Ravallion. 2001. "Household Income Dynamics in Rural China". *Policy Research Working Paper*, No. 2706. Washington DC: World Bank.

[14] Knight, John and Linda Yueh, 2002. "The Role of Social Capital in the Labor Market in China". *Department of Economics Discussion Paper*. Oxford University.

[15] Knight, John and Shi Li. 1997. "Cumulative Causation and Inequality among Villages in China". *Oxford Development Studies*, 25(2): 149–172.

[16] Li, Guo and Xiaobo Zhang. 2003. "Does *Guanxi* Matter to Nonfarm Employment?". *Journal of Comparative Economics*, 31(2): 315–331.

[17] Lin, Nan. 2001. *A Theory of Social Structure and Action*. Cambridge: Cambridge University Press.

[18] Li, Hongbin, Pak W. Liu, Ning Ma and Junsen Zhang. 2007. "Economic Returns to Communist Party Membership: Evidence from Urban Chinese Twins". *The Economic Journal*, 117(523): 1504–1520.

[19] Li, Shuang, Ming Lu and Hiroshi Sato. 2008. "The Value of Power in China: How Do Party Membership and Social Networks Affect Pay in Different Ownership Sectors?" (in Chinese) *World Economic Papers*, forthcoming.

[20] Lu, Ming, Hiroshi Sato and Shuang Zhang. 2008. "Does Social Capital Help Households Cope with Natural Disaster during Marketization?—Evidence from Rural China". Working Paper, Fudan University.

[21] Luke, Nancy and Kaivan Munshi. 2006. |New Roles for Marriage in Urban Africa: Kinship Networks and the Labor Market in Kenya". *Review of Economics and Statistics*, 88(2): 264–282.

[22] Montgomery, James D. 1991. "Social Networks and Labor Market Outcomes: Towards an Economic Analysis". *American Economic Review*, 81(5): 1408–1418.

[23] Morduch, Jonathan and Terry Sicular. 2002. "Rethinking Inequality Decomposition—With Evidence from Rural China". *The Economic Journal*, 112(1): 93–106.

[24] Morduch, Jonathan and Terry Sicular. 2000. "Politics, Growth, and Inequality in Rural China: Does It Pay to Join the Party?" *Journal of Public Economics*, 77(3): 331–356.

[25] Munshi, Kaivan and Mark Rosenzweig. 2006. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy". *American Economic Review*, 96(4): 1225–1252.

[26] Munshi, Kaivan. 2003. "Networks in Modern Economy: Mexican Migrants in the US Labor Market". *The Quarterly Journal of Economics*, 118(2): 549–599.

[27] Narayan, Deepa and Lant Pritchett. 1997. "Cents and Sociability: Household Income and Social Capital in Rural Tanzania". *Economic Development and Cultural Change*, 47(4): 871-897.

[28] Oaxaca, Ronald. 1973. "Male and Female Wage Differentials in Urban Labor Market". *International Economic Review*, 14(3): 693–709.

[29] Peng, Yusheng. 2004. "Kinship Networks and Entrepreneurs in China's Transitional Economy". *The American Journal of Sociology*, 109(5): 1045–1074.

[30] Perkins, Tamara. 2003. *Village, Market, and Well-being in a Rural Chinese Township.* London and New York: Routledge.

[31] Putnam, Robert D., Robert Leonardi and Raffaella Nanetti. 1993. *Making Democracy Work: Civic Tradition in Modern Italy*. Princeton: Princeton University Press.

[32] Sato, Hiroshi. 2003. *The Growth of Market Relations in Post-reform Rural China: A Micro-Analysis of Peasants, Migrants and Peasant Entrepreneurs*, London: RoutledgeCurzon.

[33] Sato, Hiroshi. 2008. "The Impact of Village-Specific Factors on Household Income in Rural China", in Björn Gustafsson, Li Shi, and Terry Sicular (eds.) *Inequality and Public Policy in China*, New York and Cambridge: Cambridge University Press, 287–311.

[34] Shorrocks, Anthony F. 1999. "Decomposition Procedures for Distributional Analysis: a Unified Framework based on the Shapley Value". Mimeo. Department of Economics. University of Essex.

[35] Stiglitz, Joseph E. 2000. "Formal and Informal Institution", in *Social Capital: A Multi-Faceted Perspective* (edited by Partha Dasgupta and Ismail Serageldin). Washington DC: World Bank.

[36] Stiglitz, Joseph E., Athar Hussain and Nicholas Stern. 2000. "Chinese Reforms from a Comparative Perspective", in *Incentives, Organization, and Public Economics: Papers in Honor of Sir James Mirrlees* (edited by Peter J. Hammond and Gareth D. Myles). Oxford: Oxford University Press.

[37] Walder, Andrew G. 2002. "Income Determination and Market Opportunity in Rural China, 1978–1996". *Journal of Comparative Economics*, 30(2): 354–375.

[38] Wan, Guanghua. 2004. "Accounting for Income Inequality in Rural China: A Regression-based Approach". *Journal of Comparative Economics*, 32(2): 348–363.

[39] Wan, Guanghua and Zhangyue Zhou. 2005. "Income Inequality in Rural China: Regression-based Decomposition Using Household Data". *Review of Development Economics*, 9(1): 107–120.

[40] Wan, Guanghua, Ming Lu and Zhao Chen. 2006. "The Inequality–Growth Nexus in Short and Long Run: Empirical Evidence from Rural China". *Journal of Comparative Economics*, 34(4): 654–667.

[41] Wan, Guanghua and Yin Zhang. 2006. "The Impact of Growth and Inequality on Rural Poverty in China". *Journal of Comparative Economics*, 34(4): 694–712.

[42] Wan, Guanghua, Ming Lu and Zhao Chen. 2007. "Globalization and Regional Income Inequality: Empirical Evidence from within China". *Review of Income and Wealth*, 53(1): 35–59.

[43] Wang, Xiaolu. 2006. *Income Inequality and its Influencing Factors*. Working Paper No. 126. UNU-WIDER.

[44] Wooldridge, Jeffrey M. 2003. *Introductory Econometrics, A Modern Approach*. Mason: Southwestern College Publications.

[45] Xu, Qing, Shichao Tian, Zhigang Xu, Ting Shao. 2008. "Land Institution, the Finely-Divided Land and Income Inequality in Rural China", *Economic Research Journal* (in Chinese), 2: 83–92.

[46] Yan, Yunxiang. 1996. *The Flow of Gifts: Reciprocity and Social Networks in a Chinese Village.* Palo Alto: Stanford University Press.

[47] Yue, Ximing, Terry Sicular, Li Shi, and Björn Gustafsson. 2008. "Explaining Incomes and Inequality in China", in Gustafsson, Björn, Li Shi, and Terry Sicular (eds.), *Inequality and Public Policy in China*, New York and Cambridge: Cambridge University Press, 88–117.

[48] Zhang, Shuang, Ming Lu, Yuan Zhang. 2007. "Does the Effect of Social Capital on Poverty Reduction Fall or Rise During Marketization? Evidence from Rural China". *China Economic Quarterly* (in Chinese), 6(2): 539–560.

Appendix

In an effort to test the necessity of grouping our data into Eastern areas and Middle–Western areas, we use an F-test for these groups.

First, we construct a restricted model; i.e., simply including the area dummy and other income determination variables. The model follows:

$$Lny_{ij} = a_0 + \beta_1 AR + \beta_2 SN_{ij} + \beta_3 FC_{ij} + \beta_4 OC_{ij} + \beta_5 HC_{ij} + \beta_6 PC_{ij} + \beta_n FE_j + u$$

where *AR* is the dummy variable and takes a value of 1 for Eastern areas and 0 for Middle–Western areas. Then we acquire the SSR_r, which is about 493.3. We then construct the unrestricted model, meaning that there is an extra interaction term between area and the other variables, as follows:

$$Lny_{ij} = a_0 + \beta_1 AR + \beta_2 SN_{ij} + \beta_3 AR*Var + \beta_4 FC_{ij} + \beta_5 OC_{ij} + \beta_6 HC_{ij} + \beta_7 PC_{ij} + \beta_n FE_j + u$$

where *Var* includes SN, FC, OC, HC and PC, and *AR*Var* denotes the interaction term between area and the other variables. The *SSRur* in this model is about 484.1.

We define the F-statistic as:

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

where q is the incremental number of variables involved in moving from the restricted model to the unrestricted model (11 here) and $n-k-1$ is the degrees of freedom of the unrestricted model (856). We then calculate the F-statistic, which is about 1.64, suggesting there is a systematic difference between the coefficients of the two groups.

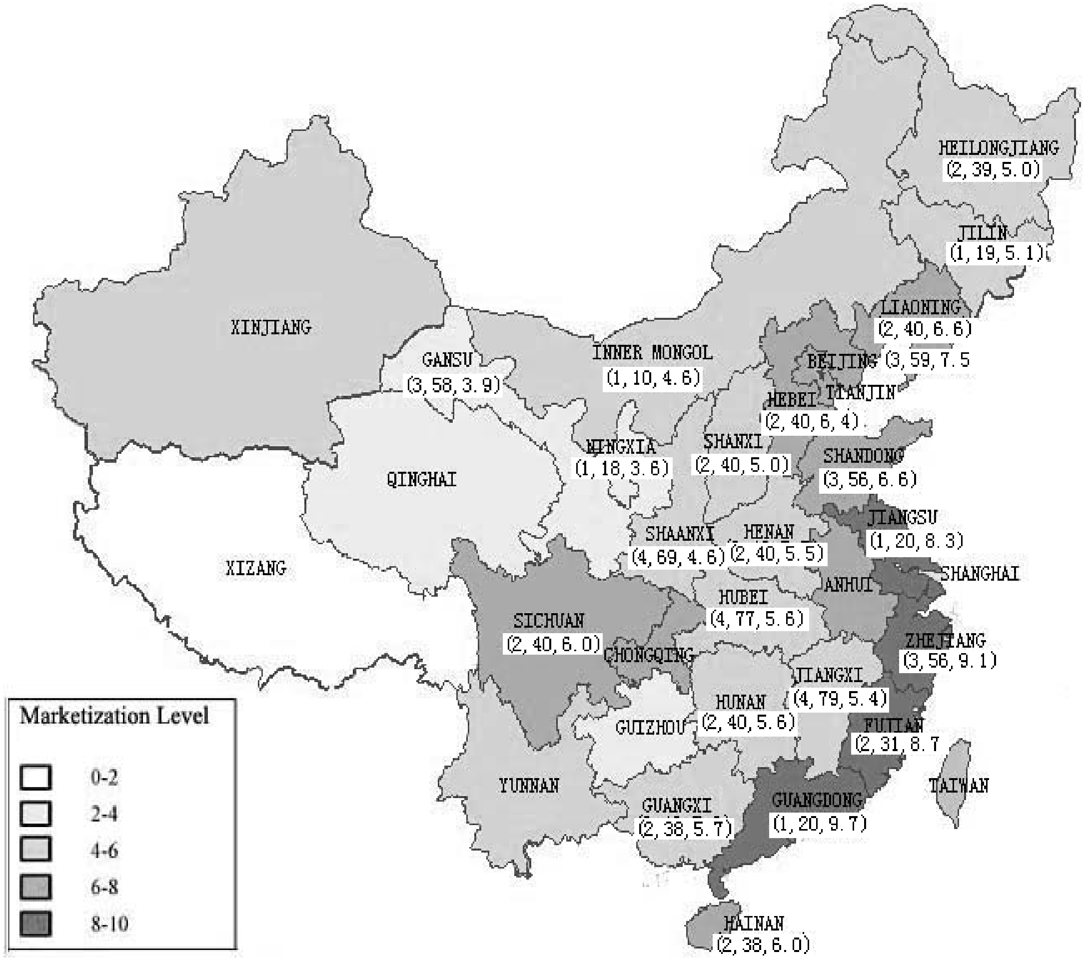

Figure 1: Sample Distribution and Marketization Level by Province in China

Notes: [1] Our sample includes 22 provinces with parentheses below.

[2] In each parenthesis following the name of a province, the first figure is the number of villages, the second is the number of households, and the third is the marketization level.

[3] Darker shading reflects higher levels of marketization. The classification used for the marketization level is in the box at the bottom-left side of the figure.

Table 1: Variable Explanation

Categories	Variables	Explanation
SN (social network)	sumrelat	the number of relatives and close friends in city or local government
	ratiosne	the ratio of gift expenditure to total daily expenditures
FC (family characteristics)	numbpeop	the number of family members
	workerave	workers per capita
	ratmale	the ratio of male workers
	ratnonaw	the ratio of nonagricultural workers
OC (physical capital)	landpercap	land per capita
HC (human capital)	worage	the average age of workers
	workeduav	the average years of education of workers
	workeduas	the square of average education
PC (political capital)	partmave	the ratio of party members
FE (fixed effect)	FE	village dummy variables
Dependent variable	lny	the logarithm of income per capita

Table 2: The Statistical Description of Variables

Categories	Variables	Total N = 927 Mean	Total N = 927 S.E.	Middle–Western N = 567 Mean	Middle–Western N = 567 S.E.	Eastern N = 360 Mean	Eastern N = 360 S.E.
SN	sumrelat	2.21	3.05	2.20	3.34	2.26	2.53
	ratiosne	0.48	0.27	0.52	0.26	0.43	0.27
FC	numbpeop	3.98	1.32	4.10	1.31	3.79	1.32
	workerave	0.81	0.20	0.79	0.19	0.82	0.20
	ratmale	0.74	0.42	0.78	0.48	0.68	0.32
	ratnonaw	0.35	0.43	0.32	0.41	0.39	0.44
OC	landpercap	1.60	2.65	1.48	2.51	1.80	2.83
HC	worage	50.02	89.59	50.65	100.71	49.03	68.62
	workeduav	7.22	4.83	6.93	4.99	7.69	4.51
	workeduas	75.47	121.06	72.98	128.99	79.38	107.45
PC	partmave	0.06	0.14	0.06	0.14	0.06	0.15
Dependent Variable	Ln y	7.86	1.06	7.61	0.99	8.26	1.46

Table 3: The Regression Results for the Total Sample

Variables		Regression Results			
		Model (1)		Model (2)	
Category	Variables	Coefficient	S.E.	Coefficient	S.E.
SN	sumrelat			0.057***	0.01
	ratiosne			0.372***	0.144
	numbpeop	0.039	0.025	0.043*	0.024
FC	workerave	0.821***	0.159	0.756***	0.156
	ratmale	0.285***	0.085	0.297***	0.082
	ratnonaw	0.618***	0.078	0.573***	0.078
OC	landpercap	0.029**	0.014	0.032***	0.014
HC	worage	0.0001	0.0003	0.0001	0.0003
	workeduav	0.127***	0.015	0.118***	0.014
	workeduas	–0.003***	0.0005	–0.003***	0.0005
PC	partmave	0.363*	0.192	0.155	0.190
Constant	a_0	6.315***	0.286	6.159***	0.286
FE	village dummies	YES			
Adjusted R^2		0.47		0.50	
No. of Obs.		927		927	

Notes: [1] Model (1) and (2) are all regressed on village dummies, whose coefficients and standard errors are omitted.

[2] *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 4: The Percentage of Explained Income Inequality

Data Group	Index	Value		Percentage of Explained Inequality
		Total Value	Explanation	
Total	Gini coefficient	0.55	0.44	80%
	Atkinson index	0.44	0.28	64%
	GE_0	0.58	0.33	57%
Middle–Western	Gini coefficient	0.48	0.40	83%
	Atkinson index	0.36	0.23	64%
	GE_0	0.49	0.27	55%
Eastern	Gini coefficient	0.57	0.43	75%
	Atkinson index	0.45	0.26	58%
	GE_0	0.60	0.31	52%

Table 5: The Decomposition Results for the Total Sample

Variables	Contribution (%)			Ranking		
	Gini	Atkinson	GE_0	Gini	Atkinson	GE_0
village dummies	38.5	43.2	42.9	1	1	1
ratnonaw	20.5	22.5	22.9	2	2	2
education	18.4	18.3	18.5	3	3	3
SN: total	13.4	12.1	12.1	4	4	4
sumrelat	9.6	9.6	9.6			
ratiosne	3.8	2.5	2.5			
ratmale	5.8	5.2	5.3	5	5	5
numbpeop and workerave	1.8	–1.3	–1.7	6	9	9
landpercap	0.8	–0.7	–0.8	7	8	8
partmave	0.6	0.6	0.6	8	6	6
worage	0.2	0.2	0.2	9	7	7
Total	100	100	100			

Table 6: The Regression Results after Grouping

Variables		Regression Results			
		Middle–Western		Eastern	
Category	Variables	Coefficient	S.E.	Coefficient	S.E.
SN	sumrelat	0.052***	0.012	0.063***	0.020
	ratiosne	0.268	0.171	0.636**	0.275
	numbpeop	–0.020	0.030	0.130***	0.040
FC	workerave	0.522***	0.197	1.161***	0.256
	ratmale	0.247***	0.095	0.368**	0.178
	ratnonaw	0.630***	0.094	0.492***	0.133
OC	landpercap	0.033**	0.016	0.037**	0.030
HC	worage	0.0003	0.0003	–0.0005	0.0006
	workeduav	0.096***	0.017	0.140***	0.027
	workeduas	–0.003***	0.0006	–0.003***	0.0001
PC	partmave	–0.18	0.245	0.397	0.304
Constant	cons	6.794***	0.333	5.02***	0.404
FE	village dummies	Yes			
Adjusted R^2		0.45		0.48	
No. of Obs.		360		567	

Notes: [1] Models (1) and (2) are all regressed on the village dummies, whose coefficients
and standard errors are omitted.
[2] *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 7: The Decomposition Results after Grouping

Variables	Contribution (%)						Ranking					
	Gini		Atkinson		GE_0		Gini		Atkinson		GE_0	
	M–E	E	M–E	E	M–E	E	M–E	E	M–E	E	M–E	E
village dummies	37.5	26.4	40.4	28.0	40.1	28.0	1	1	1	2	1	2
ratnonaw	23.7	18.4	26.1	21.2	26.4	21.7	2	3	2	3	2	3
education	13.5	25.7	11.6	29.8	11.6	29.9	3	2	4	1	4	1
SN: total	12.9	17.8	12.8	16.7	12.8	17.9	4	4	3	4	3	4
sumrelat	10.2	9.3	11.0	8.8	11.0	8.9						
ratiosne	2.7	8.5	1.8	7.9	1.8	8.0						
ratmale	7.5	5.5	7.5	5.4	7.7	5.6	5	5	5	5	5	5
numbpeop and workerave	2.9	4.5	1.2	0.03	1.1	–0.8	6	6	6	7	6	8
landpercap	1.2	0.2	–0.05	–2.0	–0.2	–2.3	7	8	8	9	9	9
partmave	–0.04	1.6	–0.09	1.3	–0.09	1.4	9	7	9	6	8	6
worage	0.8	–0.02	0.7	–0.34	0.7	–0.38	8	9	7	8	7	7
Total	100	100	100	100	100	100						