Analysis of Stochastic Dominance Ranking of Chinese Income Distributions by Household Attributes

Esfandiar Maasoumi[1], Biwei SU[2] and Almas HESHMATI[2♣]

[1] Department of Economics, Emory University, Atlanta, USA
E-mail: esfandiar.maasoumi@emory.edu

[2] Department of Food and Resource Economics
College of Life Science and Biotechnology, Korea University,
East Building Room 217, Anam-dong, Seongbuk-gu,
Seoul 136-713, Korea
E-mails: sue925sky@korea.ac.kr and heshmati@korea.ac.kr

January 17, 2013

Abstract

In this paper, we employ stochastic dominance analysis on Chinese Household Nutrition Survey (CHNS) data to investigate the inequality and relative welfare levels in China over time and among population subgroups. We find that from the period of 2000 to 2009, welfare has been continuously improved along with Chinese economic development and growth. Our pairwise comparison of population subgroups shows that there is no dominance relation between subgroups for household type, gender of households head, and age cohorts. While married group and non-child rearing group second order dominate single/divorced group and child rearing group, showing higher level of welfare in the former groups. Also, we find inequality in subgroups with different educational levels and household sizes that the groups with a higher level of education and smaller size of household tend to be better off than their counterparts.

JEL Classification Number: C14; D31; D63; O15;

Key words: Income Distribution; Stochastic Dominance; Welfare ranking; Nonparametric Testing; China;

♣ Corresponding author

1. Introduction

Improvement of a social welfare is often based on complete order ranking of income profiles in respect to various attributes. The assessment of well-being in terms of income inequality is traditionally captured by Lorenz curve and Gini coefficient. The Lorenz curve plots the cumulative share of total income against the cumulative proportion of income receiving units. Its divergence from a line of perfect equality can be measured by several indices of inequality, among which one of the most widely used is the Gini coefficient (Heshmati, 2006). The gain from such approaches is that they produce complete strong rankings of income distributions. However, they lack broad acceptance of their underlying value judgment: the attained conclusions based on these approaches often vary with the choice of types of indices and the parameters used to compute those indices. This ambiguity is problematical for policy analysis and decision making. In addition, indices are summary measures and therefore, often fail to reveal the finer details at the whole income distribution and for different subgroups which is inadequate for providing policy recommendations (Maasoumi and Heshmati, 2000). As a result, an increasing number of researches have been carried out using approaches such as stochastic dominance (SD) to avoid the above mentioned problems in recent years. Based on the expected utility paradigm, this approach is weaker than the former ones, yet provides uniform assessment of distributions between pairwise population subgroups.

Stochastic dominance relations are defined over relatively large classes of utility functions and have the potential to offer powerful majority inferences in regard with the welfare rankings of different distributions. Moreover, recently developed non-parametric tests enable one to assess such relations to a degree of statistical certainty (Millimet and Wang, 2006). Unlike average outcome from complete order ranking, which mask the differential impact of factors on different observations, stochastic dominance considers the whole distribution and reveals all of the distributional changes in incomes amongst subgroups. If dominance relations are inferred, one can distinguish the better off group from its counterparts and therefore have a clear vision of the future policies, strategies and programs for helping the targeted groups. On the other hand, the inability to infer a dominance relation is equally valuable and informative; indicating that any welfare ordering based on a particular index is highly subjective and does not apply to all segments of population.

The power of such stochastic dominance relations has led to their growing application in analysis of income inequality and well-being within or between different regions. For example, Massoumi and Heshmati (2000) analyze changes in the Swedish income distribution over time as well as across different population subgroups. Similar approaches have also been applied in another paper studying about dominance ranking of US household income by various household attributes in a period of 1968 to 1997 (Massoumi and Heshmati, 2008). Lean and Valenzuela (2012) employ stochastic dominance analysis on Australian income and expenditure distributions for the population to study inequality and relative welfare in Australia over the period 1983 to 2004. Sarkar (2012) examine the performance of rural India, urban India, female headed household and backward caste household in terms of poverty, inequality and welfare for the reference periods of 2009-2010 and 2004-2005.

Surprisingly, in case of China, stochastic dominance has been seldom applied. Existed studies generally use three kinds of approaches to study income inequality between urban and rural regions, genders and provinces: (i) statistical analysis of the range and variance of income (e.g. Sicular, et al, 2007), (ii) regression-based methods and Blinder-Oaxaca decomposition (e.g. Shi, et al., 2002), and (iii) Gini and Theil's inequality indices (e.g. Wu and Perloff, 2005; Kanbur and Zhang, 2005; Beonnefond and Clement, 2012). To our knowledge, three of the studies use stochastic dominance technique for analysis of inequality or welfare in China. One is from Millimet and Wang (2006). The authors utilize recent developed nonparametric test for stochastic dominance to assess and decompose the distribution of gender earnings differentials in urban China in the mid-1990s. They find large earnings differentials across gender in the lower tail of the distribution using both annual earnings and hourly wage, but no differential in the upper tail. The second research is from Anderson and Ge (2009), in which they find evidence of strong welfare gains from 1990 to 1999 in all regions together with significant and persistent welfare disparities between cities within Eastern coastal area and those in the interior. And the last one is from Cai, et al. (2010), which is about changes in income and consumption inequality in China's urban population during the period 1992-2003. In order to contribute to the existing literatures, we examine the dominance relations for more recent years, from 2000 to 2009 and for several household attributes that have not been studied yet.

In this paper, we consider statistical test procedures for first and second order stochastic dominance. The tests studied here are multivariate generalizations of the Kolmogorov-Smirnov statistics when weak dependence is permitted in the processes. Four waves of household survey data have been used for analysis in two parts. Part one is "unconditional" tests for stochastic dominance between years for the entire distribution of the annual average incomes. Part two contains "conditional" tests for stochastic dominance between population subgroups which are categorized by household attributes including household registration type, household size, age, gender, education, marital and child rearing status. Here, "conditional" implies before stochastic dominance tests for subgroups of one certain household attribute, say household registration type, we control for influences from all other household attributes that mentioned above through regression analysis and data transformation.

This study proceeds as follows. Section 2 introduces the data that we used in this research. Section 3 provides theory, methods as well as the test statistics. Section 4 gives results analysis. Section 5 contains the summary of the major findings and conclusion.

2. Data and Data Transformation

In this study, we use China Health and Nutrition Survey (CHNS) household data which is collected by the Carolina Population Center at the University of North Carolina at Chapel Hill, the Institute of Nutrition and Food Hygiene, and the Chinese Academy of Preventive Medicine[1]. This on-going longitudinal survey has in total eight waves: 1989, 1991, 1993, 1997, 2000, 2004, 2006, and 2009. The sample households are randomly drawn from nine provinces including Heilongjiang, Liaoning, Shandong, Jiangsu, Henan, Hubei, Hunan, Guangxi, and Guizhou (Figure 1 shows the map of the regions) each wave covering about 3,800 households with a total of 14,000 individuals. While

[1] A detailed description of the data and quality control procedures can be obtained from http://www.cpc.unc.edu/projects/china.

we trace the surveyed households over time, the sample size drops rapidly. Finally, we form a panel data set of 1,225 households from four waves of the survey: 2000, 2004, 2006 and 2009 covering information of household per capita income and social or demographic characteristics for each household.

The darker shaded regions in this map are the provinces in which the survey

Figure 1. Map of CHNS Survey Regions

Household per capita income accounts for all forms of cash and in-kind income, including the value of farm products produced by rural households and retained for self-consumption, in-kind subsidies, transfers, and earnings from assets such as rental revenues from real estate. In order to avoid biases from inflation, all the income values are transformed to fixed 2009 prices using the (consumer) price indices provided by CHNS.

The household characteristics that we control for are household registration type, household size, educational level, age, gender, marital and child rearing status of head of the household. Since several of the household characteristics change over time, we choose to use those records of head of the household head in 2009 as a reference.

Definition of income and subgroups of household are given in Table 1. Household registration type is categorized into two urban and rural groups in accordance with the birth place of the household head. In defining household size group, the sample is divided into five groups: live alone, two family members, three family members, four family members and five or above five family members. Four educational levels are included: no education, primary education, junior middle school, senior middle school and technical school education and above. Taking account of productivity, we separate the sample into three age cohorts: aged from 15 to 40, 41 to 65 and above 65. Gender groups stand for female headed households and male headed households. Marital status contains two groups, for those who are single, divorced or widow and those who are married. And last, child rearing divides the sample into two subgroups: households with a child that required daily care and the otherwise ones.

[Insert Table 1 here]

As mentioned before, our analysis is carried out in two parts: the stochastic dominance tests are conducted between pairwise groups in respect with time (unconditional) and household attributes (conditional). When observe dominance relationships over years, we compare the household per capita income in the year of 2000, 2004, 2006 and 2009. When it comes to comparison between subgroups of the household, certain data transformation is needed. We first obtain the average household per capita income over the period 2000 to 2009 as a preparation for data transformation. This is because using the average value takes out some transitory movements which otherwise cannot be achieved by using snapshots at points in time (Massoumi and Heshmati, 2008). And then we conduct the Mincerian (Mincer, 1994) OLS estimation:

$$\ln(\text{Income}) = \text{f}(\text{Province, Household registration, Household size, Gender, Education, Agecohort, Marital and child rearing status})$$

where income is the average household per capita income over the period of 2000 to 2009 and household head attributes recorded in 2009. Dummies for provinces are also added for controlling of regional heterogeneous. Last, we subtract the effects of controlled characteristics except for one specific characteristic (such as household registration type shown below for illustration) from average income values. And the remains are the controlled income that we use for SD test.

$$\text{Residuals} + \text{Registration type's effects} = \ln(\text{income}) - \text{Other characteristics' effects } (\text{Xi} * \beta\text{i})$$

3. Theory, methods and the test statistics

In lighted by Massoumi and Heshmati's (2008) work on ranking PSID incomes by various household attributes, we use similar bootstrap procedure for estimating the probability of rejection of the SD hypotheses with a suitably extended Kolmogorov-Smirnov test for first and second order stochastic dominance.

3.1 Theory and methods

Let X and Y be two income variables at either two different points in time, or for different regions, or individual or household characteristics. Let $X_1, X_2, \ldots, X_n$, be n not necessarily i.i.d. on X, and $Y_1, Y_2, \ldots, Y_n$, be n not necessarily i.i.d. on Y income variables. Let U_1 denote the class of all von-Neumann-Morgenstern-type utility functions (Von-Neumann, 1953), u, such that $u' \geq 0$, (increasing). Also, let U_2 denote the class of all utility functions in U_1 for which $u'' \leq 0$ (strict concavity), and U_3 denote a subset of U_2 for which $u''' \geq 0$. Let F(x) and G(x) denote the cumulative distribution functions for the two comparison groups, respectively. Quantiles $q_x(p)$ and $q_y(p)$ are implicitly defined by, for example, $F[X \leq q_x(p)] = p$.

Definition 1. X First Order Stochastic Dominants Y, denoted X FSD Y, if and only if any one of the following equivalent conditions holds:

(1) $E[u(X)] \geq E[u(Y)]$ for all $u \in U_1$, with strict inequality for some u.

(2) $F(x) \leq G(x)$ for all x in the support of X, with strict inequality for some x.
(3) $q_x(p) \geq q_y(p)$ for all $0 \leq p \leq 1$, with strict inequality for some p.

Definition 2. X Second Order Stochastic Dominant Y, denoted X SSD, if and only if any one of the following equivalent conditions holds:

(1) E [u(X)] ≥ E [u(Y)] for all u $\in U_2$, with strict inequality for some u.

(2) $\int_{-\infty}^{x} F(t)dt \leq \int_{-\infty}^{x} G(t)dt$, for all x in the support of X and Y, with strict inequality for some x.

(3) $\int_{0}^{p} q_x(t)dt \leq \int_{0}^{p} q_y(t)dt$, for all $0 \leq p \leq 1$, with strict inequality for some p.

The tests of FSD and SSD are based on empirical evaluations of conditions (2) or (3) in the above definitions. Mounting test on condition (2) requires empirical commutative density function (*cdf*s) and comparisons at a finite number of observed ordinates. Mounting tests on conditions (3) typically relies on the fact that quantiles are consistently estimated by the corresponding order statistics at a finite number of sample points.

This approach fixes the critical value (zero) at the boundary of our null and estimates the associated "significance level" by bootstrapping the sample or its blocks. This renders our tests "asymptotically similar" and unbiased on the boundary which is similar in spirit to inference based on p-values.

3.2 The test statistics

Suppose that there are two prospects X_1 and X_2 and let A= $\{X_k$: k=1,2$\}$. Let $\{X_{ki}$:i=1,2…N$\}$ be realizations of X_k for k=1,2. Here we group the data into subgroups, say of household registration type, or household heads gender, and then make comparisons across homogenous populations. For k=1,2 define:

$F_k(x,\theta) = P(X_{ki}(\theta) \leq x)$

and

$$\overline{F}_{kN}(x,0) = \frac{1}{N}\sum_{i=0}^{N} 1(X_{ki}(0) \leq x)$$

We denote $F_k(x) = F_k(x, 0_{k0})$ and $\overline{F}_{kN}(x, 0_{k0})$ and let F (x_1, x_2) be the joint c.d.f. of $(X_1, X_2)'$. Now define the following functional of the joint distribution:

$$d = \min_{k \neq l} \sup_{x \in \chi} [F_k(x) - F_l(x)]$$

$$s = \min_{k \neq l} \sup_{x \in \chi} \int_{-\infty}^{x} [F_k(x) - F_l(x)]dt$$

where χ denotes a given set contained in the union of the supports of X_{ki} for k=1,2. Without loss of generality we assume that the supports are bounded. The hypotheses of interest are:

$$H_0^d : d \leq 0 \ vs. \ H_1^d : d > 0$$

$$H_0^s : s \leq 0 \ vs. \ H_1^s : s > 0$$

The null hypothesis H_0^d implies that the prospects in A are not first-degree stochastically maxima, i.e., there exists at least one prospect in A which first-degree dominates the others. Likewise for the second order case, the test statistics are given below:

$$D_N = \min_{k \neq l} \sup_{x \in \chi} \sqrt{N}\, [\overline{F}_{kN}(x, \dot{0}_k) - \overline{F}_{lN}(x, \dot{0}_l)]$$

$$S_N = \min_{k \neq l} \sup_{x \in \chi} \sqrt{N} \int_{-\infty}^{x} [\overline{F}_{kN}(x, \dot{\theta}_k) - \overline{F}_{lN}(x, \dot{\theta}_l)]$$

In our algorithm we compute approximation to the suprema in D_N, S_N based on taking maxima over some smaller grid of points $X_J = \{x_1, \dots, x_j\}$ where $J < n$. We obtain simple bootstrap estimates of the probability $\{D_N \leq 0\}$ and probability$\{S_N \leq 0\}$. If the probability is high, say 0.90 or higher, we may infer dominance to a desirable degree of confidents. If it is a low probability, say 0.10 or smaller, we may infer the presence of significant crossing of the empirical CDFs, implying an inability to rank the outcomes (Massoumi and Heshmati, 2008).

4. Analysis of the results

4.1 Unconditional analysis

The first part of Table 1 shows the summary of the data by years of observation. The balance number of households observed all four waves is 1,225. The mean income is continuously increasing over time from 5,606 yuan[2] in 2000 to 13,410 yuan in 2009, that is, even after adjusting for inflation, the income has been doubled in 10 years. Similarly, the dispersion in income increases from 5,651 yuan in 2000 to 13,388 yuan in 2009.

[Insert Table 2 here]

The second part of Table 2, our SD test statistics are summarized by their mean and standard errors, as well as the probability of the test statistic. Subgroups are compared pair wisely, one group is denoted as the "X" distribution and the other group is denoted by "Y". Thus, "FSDxoy" denotes "first order stochastic dominance of X over Y", and "SSDxoy" is similarly is defined for second order dominance. "FOmax" and "SOmax" denotes the "first and second order maximality".

Following the statistic test procedure, the test shows that there is no FSD between any two years. The probability for first order stochastic dominance varies from 0.0000

[2] Chinese currency $1=6.23 yuan on December 29, 2012.

to 0.3170, suggesting that the null hypothesis --- that there is first order stochastic dominance between two groups--- is unlikely to be accepted. However, the second stochastic dominance relationship has been observed in all comparisons and without exception, the latter years SSD the prior years. Thus those results imply that an increase in social welfare over recent 10 years in China regardless of which social welfare function or index we use for measurement.

4.2 Conditional analysis

As mentioned previously, the households are distinguished by household heads characteristics in 2009, but mean income is defined as period average of household per capita income. The characteristics that we condition on include: household registration type, gender age, household size, education, marital and child rearing status. The conditional analysis of the data for each characteristic is discussed below.

Test Results for household registration type

The mean values for household registration type reported in Table 3 gives a higher average income for urban group (10,771yuan) than that for rural group (8,878yuan). Surprisingly, the results from the SD test reveal that there is neither first nor second dominance of the urban households over the rural households. Since the probability for urban households FSD rural households is 0.53, and that for SSD is 0.68 --- both are weak evidence of stochastic dominance. This raise contrast with previous studies (Khan and Riskin, 1998; Benjamin, et al., 2005; Heshmati, 2007, among others) which indicate significant rural-urban inequality in China. We consider this might be because the attributes that we controlled for are also attributed to urban-rural inequality to certain extent, for example urban residents get more opportunities and better quality of education than rural residents. Thus after eliminating those effects, the inequality between urban and rural residents is not as obvious as before.

[Insert Table 3 here]

Test Results for gender of households head

From Table 3 we can see that male headed households are nearly 10 times as many as female headed households suggesting that China is still a male-dominated country. Even though on average female headed households (FHH) have higher incomes than male headed households (MHH), there is no evidence of FHH first or second order stochastic dominating over MHH. This basically implies that other factors being equal, FHH and MHH enjoy the same level of well-being. At a first glance, one might note that this finding runs in contrast to the finding of gender inequality in China's society in the earlier works (Macpherson and Hirsch, 1995; Hughes and Maurer-Fazio, 2002; Wang and Cai, 2006 among others). A careful comparison will however, reveals that previous studies use individual income data and do not control for individual's attributes. Also, in a male dominated society, female heads is likely to work harder in order to compete with their counterparts.

Test results for marital and child rearing status

In Table 3, it is found that one tenth of household heads is single or divorced; the mean value of single/divorced group income is lower than that for the married one, which is 9,137 yuan vs. 10,006 yuan. This difference has also been proved through the SD test, suggesting that the married group second order dominates the single/divorced group at 91 per cent level of significance. Around 90% of the households indicate that they do not spend time on rearing child at the survey year. The mean and the dispersion of income for that group are higher than those for households that need to spend time for children rearing. In the SD test, the former group is second order dominating the latter one at 96 per cent level showing that taking care for a child may have negative influence on household's welfare level.

Test results for age cohorts

Table 4 summarizes the results for age group 15-40, 41-65, and 65+. The mean values and the dispersion of incomes for the three age cohorts have little differences from one to another. The group that represent household head's age between 40 and 65 has the highest income level but the lowest dispersion. Looking at the SD test results, there are no cases for FSD or SSD between the age cohorts. The disposable incomes of age cohorts are second order maximal, implying that they may be ranked only at higher levels than SSD. The results reveal that though age is considered highly related to one's productivity and therefore one's income level, this is less true from seeing at household level. Since chances are high that some low income earner can be compensated by financial supports from other family members who have higher income level.

[Insert Table 4 here]

Test results for educational levels

It is generally held that education has positive effect on earnings and bargaining power. From Table 5, we can see that mean incomes are increasing function of the years of schooling. Educational returns reach the highest level for people who finished technical school or tertiary education. In the SD test, we find that the group for technical/tertiary education SSD the groups of no education, primary and junior high school education. Nevertheless, the group for technical/tertiary education shows no dominance relationship over the group for senior high school education, even if the mean income difference between those two groups is as large as 4,866 yuan. The group for senior high school education SSD the groups for no education and primary education, but no ranking order can be established while comparing with the group for junior high school education. For the rest of the pair comparisons, only the group of junior high school education indicates a SSD relationship with the group of no education. It seems that the comparisons between two adjacent educational levels end up with low probability, which is a weak evidence for FSD or SSD. However, SSD relationships are invariably observed when two comparing educational levels have at least one interval level in between.

[Insert Table 5 here]

Test results for household sizes

Summary of the mean income in Table 6 reflects negative effects of household size on the per capita household income level that it drops from 9,461 yuan for households with only one person (HH1) to 4,632 yuan for household with 5 or more people (HH5). The largest two groups are households with two or three family members --- the combined observations in those two groups standing for more than half of the total number of households. Yet, despite they are the two most common family patterns in Chinese society, the mean value of incomes remain below those for HH1. In terms of the SD test, the group for those who live alone (HH1) SSD the groups of households with 3 (HH3) or 4+ (HH5) family members. HH1 also second order dominates the group for household with 2 family members (HH2) but with only a probability of 0.8960. Nevertheless, it is interesting to find that there is no dominance relation between HH1 and HH4. HH2 shows a strong second stochastic dominance over HH3, HH4, and HH5. But for the comparisons between HH3, HH4 and HH5, the scale of probability is too weak to show any clear dominance ranking order relationship.

[Insert Table 6 here]

5. Summary and conclusion

Income inequality and welfare in China has long been studied at home and abroad while the extent of inequality mostly been measured and ranked using partial strong orders such as the Gini coefficient or the Atkinson Index. Approaches like those are known to suffer from the lack of universal acceptance of the value judgments of the underlying welfare functions which sometimes result in contradicting conclusions. The purpose of this paper is to avoid those shortcomings and draw a more robust conclusion on income inequality level in China. To achieve this, we consider Kolmogorov-Smirnov test procedures for first and second order stochastic dominance relations in respect with time and several household attributes in China.

The SD tests results show a steadily rising per capita household welfare from 2000 to 2009. This is both expected and a desirable outcome of the economic development. Since the rapid economic growth in contemporary China undeniably has raised the quality of life. Our subgroup-specific findings suggest a number of important items for policy concern. Even though first order dominance is rare, but second order dominance holds in most of cases. First, we find that after controlling for other attributes, there is no stochastic dominance relation between urban and rural residents, female-headed and male-headed household and age cohorts. That is to say that those attributes may have limited contributes to increased income inequality. Second, our results show that the group of married household heads SSD the group of single/divorced ones, and similarly the group of non-child rearing SSD the group of child-rearing, pointing that the former group has a higher level of welfare than the latter one. Third, groups with higher levels of education second order dominates groups with the lower levels of education suggesting that the latter groups have higher inequality when compared to the former groups. And fourth, large-size families have higher income inequality and lower welfare level than small-size families.

Based on our findings, we conclude that ceteris paribus influence of household registration type, gender of the household head and age cohort have less contribution to income inequality than other household attributes. Household heads who are single/divorced or have low level of education and households with a child needs for caring or with a large family number tend to have lower welfare level. Thus, policies that are favorable to those population subgroups are recommended for reduction or elimination of inequality in China.

While we believe our study, based on typical Chinese provinces, represents some interesting findings about the income inequality in China. One should be cautious that our samples only cover nine out of thirty-one provinces. There is a risk of oversimplification in attempting to summarize our key findings and apply it to China as a whole. If existed, further research could ideally be conducted using national wide household-level data set.

References

Anderson, G. and Ge, Y. (2009). Intercity Income Inequality Growth and Convergence in China. *Journal of Income Distribution* 18(1), 1-41.

Benjamin, D., Brandt, L., and Giles, J. (2005). The Evolution of Income Inequality in Rural China. *Economic Development and Cultural Change* 53(4), 769-824.

Bonnefond, C. and Clement, M. (2012). An Analysis of Income Polarisation in rural and urban China. *Post-Communist Economies* 24(1), 15-37.

Cai, H.B., Chen, Y.Y. and Zhou, L.A. (2010). Income and Consumption Inequality in Urban China. *Economic Development and Cultural Change* 58(3), 385-413.

Heshmati A. (2006), "The World Distribution of Income and Income Inequality: A Review of the Economic Literature", Journal of World Systems Research 12(1), 60-107.

Heshmati A. (2007), "Income Inequality in China", in Heshmati (Ed.), "Recent Developments in the Chinese Economy", Nova Science Publishers, NY.

Hughes, J. and Maurer-Fazio, M. (2002). Effects of Marriage, Education, and Occupation on the Female/Male Wage Gap in China. Pacific Economic Review 7(1), 137-156.

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

Khan, A.R., and Riskin, C. (1998). Income and Inequality in China: Composition, Distribution and Growth of Household Income, 1988 to 1995. *The China Quarterly* 154, 221-253.

Lean, H.H. and Valenzuela, M.R. (2012). Inequality in Australia 1983-2004: A Stochastic Dominance Approach, Department of Economics ISSN1441-5429, Discussion Paper.

Macpherson, D. and Hirsch, B. (1995) Wages and Gender Composition: Why Do Women's Jobs Pay Less? *Journal of Labor Economics* 133(3), 84-89.

Massoumi, E. and Heshmati, A. (2000). Stochastic Dominance amongst Swedish Income Distributions, *Econometric Reviews* 19(3), 287-320.

Massoumi, E. and Heshmati, A. (2008). Evaluating dominance ranking of PSID incomes by various household attributes. *Advances in Income Inequality and Concentration Measures*, Routledge, London.

Millimet, D.L. and Wang, L. (2006). A Distributional Analysis of the Gender Earnings Gap in Urban China. *Contributions to Economic Analysis and Policy* 5(1), Article 5.

Mincer, J. (1994). Schooling, Experience, and Earnings. Human Behavior and Social Institutions No. 2, National Bureau of Economic Research, NY.

Sarkar, S. (2012). Application of Stochastic Dominance: A study in India. *Procedia Economics and Finance* 1, 337-345.

Sicular, T., Yue, X.M., Gustafasson, B., and Li, S. (2007). The Urban-Rural Income Gap and Inequality in China. *Review of Income and Wealth* 53(1), 93-126.

Shi, X.Z., Sicular, T. and Zhao, Y.H. (2002). Analyzing Urban-Rural Income Inequality in China. International Symposium on Equity and Social Justice, Discussion Paper.

Tong, G.R. and Zhang, J.B. (2006). Study on the Trend of Chinese Peasant Income Distributions. *The Journal of Quantitative and Technical Economics* 8, 3-10.

Von-Neumann, J. and Morgenstern (1953). *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, N.J.

Wang, M.Y. and Cai, F. (2006). Gender Wage Differentials in China's Urban Labor Marker. UNU-WIDER, Research Paper No.146.

Wu, X.M. and Perloff, J.M. (2005). China's Income Distribution, 1985-2001. *The Review of Economics and Statistics* 87(4), 763-775.

Table 1. Description of income and subgroups of households

Variables	Definition
Income	Household per capita Income for each year/ Average Household per capita income for four year
Household registration type	Urban/ Rural
Education:	
Edu0	Received no education
Edu1	Received 6 years of education (Primary school)
Edu2	Received 9 years of education (Junior middle school)
Edu3	Received 12 years of education (Senior middle school)
Edu4	Received 14 or 14+ years of education (Technical school and above)
Age cohort:	
Agecohort1	15-40
Agecohort2	40-65
Agecohort3	66 and above
Gender of household head	Male/ Female
Marital status	Single, divorced and widow/Married
Rearing child	Yes/No
Household size:	
HHsize1	Live alone
HHsize2	2 family members in a household
HHsize3	3 family members in a household
HHsize4	4 family members in a household
HHsize5	5 or more than 5 family members in a household

Table 2. Kolmogorov-Smirnov test for stochastic dominance over years

Characteristics	**Obs**	**Mean**	**Std dev**				
2000	1,225	5,606	5,651				
2004	1,225	7,618	6,992				
2006	1,225	8,675	8,586				
2009	1,225	13,410	13,388				
Test	**Mean**	**Std err**	**Prob**		**Mean**	**Std err**	**Prob**
2000 (x) vs. 2004 (y)				2000 (x) vs. 2006 (y)			
FSDxoy	0.1594	0.0171	0.0000	FSDxoy	0.0969	0.0172	0.0000
FSDyox	0.0009	0.0012	0.0680	FSDyox	0.0176	0.0091	0.0000
FOmax	0.0009	0.0012	0.0680	FOmax	0.0176	0.0091	0.0000
SSDxoy	0.5013	0.0580	0.0000	SSDxoy	0.2831	0.0528	0.0000
SSDyox	-0.1150	0.0202	1.0000	SSDyox	0.0173	0.0097	0.0360
SOmax	-0.1150	0.0202	1.0000	SOmax	0.0173	0.0097	0.0360
2000 (x) vs. 2009 (y)				2004 (x) vs. 2006 (y)			
FSDxoy	0.3722	0.0159	0.0000	FSDxoy	0.0557	0.0153	0.0000
FSDyox	0.0000	0.0001	0.3080	FSDyox	0.0014	0.0034	0.2570
FOmax	0.0000	0.0001	0.3080	FOmax	0.0013	0.0031	0.2570
SSDxoy	0.6053	0.0335	0.0000	SSDxoy	0.2448	0.0725	0.0010
SSDyox	-0.3722	0.0159	1.0000	SSDyox	-0.0278	0.0194	0.9330
SOmax	-0.3722	0.0159	1.0000	SOmax	-0.0279	0.0193	0.9340
2004 (x) vs. 2009 (y)				2006 (x) vs. 2009 (y)			
FSDxoy	0.2552	0.0183	0.0000	FSDxoy	0.2130	0.0188	0.0000
FSDyox	0.0000	0.0000	0.3070	FSDyox	0.0000	0.0002	0.3170
FOmax	0.0000	0.0000	0.3070	FOmax	0.0000	0.0002	0.3170
SSDxoy	0.4572	0.0353	0.0000	SSDxoy	0.3792	0.0372	0.0000
SSDyox	-0.2552	0.0183	1.0000	SSDyox	-0.2130	0.0188	1.0000
SOmax	-0.2552	0.0183	1.0000	SOmax	-0.2130	0.0188	1.0000

Glossary of Tests:

FSDxoy First order stochastic dominance of **x** over y
FSDyox First order stochastic dominance of y over x
FOmax First order maximal
SSDxoy Second order stochastic dominance of **x** over y
SSDyox Second order stochastic dominance of y over x
SOmax Second order maximal
Prob Reject the null of no dominance when the statistics are negative

.

Table 3. Kolmogorov-Smirnov test for stochastic dominance relations for household registration type, gender of household head and marital and child rearing status

Characteristics	**Obs**	**Mean**	**Std dev**	**Characteristics**	**Obs**	**Mean**	**Std dev**
Rural	911	8,878	4,689	Female Headed	133	8,964	5,398
Urban	314	10,771	6,808	Male Headed	1,092	8,035	5,302
Test	**Mean**	**Std err**	**Prob**	**Test**	**Mean**	**Std err**	**Prob**
Rural (x) vs. Urban (y)				Female (x) vs. Male (y)			
FSDxoy	0.1679	0.0304	0.0000	FSDxoy	0.0198	0.0175	0.0000
FSDyox	0.0013	0.0080	0.5300	FSDyox	0.1125	0.0413	0.0000
FOmax	0.0013	0.0080	0.5300	FOmax	0.0191	0.0164	0.0000
SSDxoy	0.6551	0.1333	0.0000	SSDxoy	0.0075	0.0362	0.4600
SSDyox	-0.0074	0.0156	0.6800	SSDyox	0.2780	0.1460	0.0220
SOmax	-0.0074	0.0156	0.6800	SOmax	0.0029	0.0300	0.4820
Characteristics	**Obs**	**Mean**	**Std dev**	**Characteristics**	**Obs**	**Mean**	**Std dev**
Not rearing child	1,114	10,545	5,365	Single/divorced	130	9,137	5,028
Rearing child	111	8,814	4,730	Married	1,095	10,006	5,344
Test	**Mean**	**Std err**	**Prob**	**Test**	**Mean**	**Std err**	**Prob**
No (x) vs. Yes (y)				Single (x) vs. Married (y)			
FSDxoy	0.2634	0.0493	0.0000	FSDxoy	0.0979	0.0359	0.0000
FSDyox	0.0136	0.0085	0.0000	FSDyox	0.0127	0.0154	0.2240
FOmax	0.0136	0.0085	0.0000	FOmax	0.0122	0.0142	0.2240
SSDxoy	-0.0409	0.0219	0.9610	SSDxoy	0.3129	0.1400	0.0020
SSDyox	0.6203	0.1491	0.0000	SSDyox	-0.0463	0.0389	0.9080
SOmax	-0.0409	0.0219	0.9610	SOmax	-0.0478	0.0342	0.9100

Glossary of Tests:

FSDxoy First order stochastic dominance of **x** over y
FSDyox First order stochastic dominance of y over x
FOmax First order maximal
SSDxoy Second order stochastic dominance of **x** over y
SSDyox Second order stochastic dominance of y over x
SOmax Second order maximal
Prob Reject the null of no dominance when the statistics are negative
.

Table 4. Kolmogorov-Smirnov test for stochastic dominance relations among age cohorts

Characteristics	**Obs**	**Mean**	**Std dev**				
Age1 (15-40)	368	9,189	5,481				
Age2 (40-65)	591	9,311	5,147				
Age3 (65+)	266	9,289	5,420				
Test	**Mean**	**Std err**	**Prob**		**Mean**	**Std err**	**Prob**
Age1(x) vs. Age2 (y)				Age1 (x) vs. Age3 (y)			
FSDxoy	0.0298	0.0165	0.0090	FSDxoy	0.0465	0.0234	0.0020
FSDyox	0.0276	0.0205	0.0070	FSDyox	0.0487	0.0268	0.0000
FOmax	0.0157	0.0094	0.0160	FOmax	0.0312	0.0162	0.0020
SSDxoy	0.0707	0.0756	0.1660	SSDxoy	0.0715	0.1036	0.3270
SSDyox	0.0458	0.0666	0.3180	SSDyox	0.0958	0.0725	0.0490
SOmax	0.0038	0.0215	0.4840	SOmax	0.0159	0.0433	0.3760
Age2 (x) vs. Age3 (y)							
FSDxoy	0.0554	0.0300	0.0090				
FSDyox	0.0516	0.0284	0.0000				
FOmax	0.0326	0.0169	0.0090				
SSDxoy	0.1009	0.1332	0.3060				
SSDyox	0.1304	0.1049	0.0600				
SOmax	0.0257	0.0446	0.3660				

Glossary of Tests:

FSDxoy First order stochastic dominance of **x** over y
FSDyox First order stochastic dominance of y over x
FOmax First order maximal
SSDxoy Second order stochastic dominance of **x** over y
SSDyox Second order stochastic dominance of y over x
SOmax Second order maximal
Prob Reject the null of no dominance when the statistics are negative
.

Table 5 Kolmogorov-Smirnov test for stochastic dominance among educational levels

Characteristics	**Obs**	**Mean**	**Std dev**
Edu0 (No education)	224	6,177	4,259
Edu1 (Primary)	257	7,050	5,201
Edu2 (Junior High)	405	8,478	5,329
Edu3 (Senior High)	147	9,588	6,109
Edu4(Technical/Tertiary)	192	14,454	8,118

	Mean	**Std err**	**Prob**	**Test**	**Mean**	**Std err**	**Prob**
Edu1 (x) vs. Edu2 (y)				Edu1 (x) vs. Edu3 (y)			
FSDxoy	0.1537	0.0353	0.0000	FSDxoy	0.2346	0.0461	0.0000
FSDyox	0.0088	0.0083	0.1310	FSDyox	0.0061	0.0107	0.2530
FOmax	0.0088	0.0083	0.1310	FOmax	0.0061	0.0107	0.2530
SSDxoy	0.7005	0.1812	0.0000	SSDxoy	1.1563	0.2511	0.0000
SSDyox	-0.0145	0.0117	0.9010	SSDyox	-0.0234	0.0130	0.9570
SOmax	-0.0145	0.0117	0.9010	SOmax	-0.0234	0.0130	0.9570
Edu1 (x) vs. Edu4 (y)	**Mean**	**Std err**	**Prob**	Edu1 (x) vs. Edu0 (y)			
FSDxoy	0.4153	0.0426	0.0000	FSDxoy	0.0095	0.0127	0.1500
FSDyox	-0.0040	0.0059	0.6400	FSDyox	0.1155	0.0358	0.0000
FOmax	-0.0040	0.0059	0.6400	FOmax	0.0093	0.0121	0.1500
SSDxoy	1.6467	0.1904	0.0000	SSDxoy	-0.0192	0.0253	0.8900
SSDyox	-0.0532	0.0161	0.9990	SSDyox	0.5415	0.1945	0.0020
SOmax	-0.0532	0.0161	0.9990	SOmax	-0.0201	0.0207	0.8920
Edu2 (x) vs. Edu3 (y)				Edu2 (x) vs. Edu4 (y)			
FSDxoy	0.1188	0.0407	0.0000	FSDxoy	0.2935	0.0371	0.0000
FSDyox	0.0149	0.0150	0.1300	FSDyox	-0.0046	0.0053	0.7220
FOmax	0.0147	0.0147	0.1300	FOmax	-0.0046	0.0053	0.7220
SSDxoy	0.5280	0.2303	0.0030	SSDxoy	1.2661	0.1829	0.0000
SSDyox	0.0006	0.0262	0.7050	SSDyox	-0.0217	0.0098	0.9830
SOmax	-0.0007	0.0198	0.7080	SOmax	-0.0217	0.0098	0.9830
Edu2 (x) vs. Edu0 (y)				Edu3 (x) vs. Edu4 (y)			
FSDxoy	0.0010	0.0032	0.1880	FSDxoy	0.2260	0.0488	0.0000
FSDyox	0.2912	0.0393	0.0000	FSDyox	0.0067	0.0071	0.0600
FOmax	0.0010	0.0032	0.1880	FOmax	0.0067	0.0071	0.0600
SSDxoy	-0.0346	0.0149	0.9970	SSDxoy	0.9276	0.2015	0.0000
SSDyox	1.1739	0.1762	0.0000	SSDyox	-0.0009	0.0090	0.5070
SOmax	-0.0346	0.0149	0.9970	SOmax	-0.0009	0.0090	0.5070
Edu3 (x) vs. Edu0 (y)				Edu4 (x) vs. Edu0 (y)			
FSDxoy	0.0014	0.0036	0.0850	FSDxoy	-0.0054	0.0054	0.7330
FSDyox	0.3472	0.0479	0.0000	FSDyox	0.4977	0.0440	0.0000
FOmax	0.0014	0.0036	0.0850	FOmax	-0.0054	0.0054	0.7330
SSDxoy	-0.0424	0.0156	0.9970	SSDxoy	-0.1027	0.0213	1.0000
SSDyox	1.6042	0.2386	0.0000	SSDyox	1.8755	0.1933	0.0000
SOmax	-0.0424	0.0156	0.9970	SOmax	-0.1027	0.0213	1.0000

Glossary of Tests:

FSDxoy First order stochastic dominance of **x** over y
FSDyox First order stochastic dominance of y over x
FOmax First order maximal
SSDxoy Second order stochastic dominance of **x** over y
SSDyox Second order stochastic dominance of y over x
SOmax Second order maximal
Prob Reject the null of no dominance when the statistics are negative

Table 6 Kolmogorov-Smirnov test for stochastic dominance for household size

Characteristics	**Obs**	**Mean**	**Std dev**
HH1(Live alone)	50	9,461	6,021
HH2(2 family members)	373	7,460	6,443
HH3(3 family members)	319	6,091	5,367
HH4(4 family members)	220	5,160	3,955
HH5(4+ family members)	263	4,632	3,744

Test	**Mean**	**Std err**	**Prob**				
HH1 (x) vs.HH2 (y)				HH1 (x) vs. HH3 (y)			
FSDxoy	0.0241	0.0170	0.0010	FSDxoy	0.0139	0.0115	0.0040
FSDyox	0.2664	0.0637	0.0000	FSDyox	0.3487	0.0645	0.0000
FOmax	0.0241	0.0170	0.0010	FOmax	0.0139	0.0115	0.0040
SSDxoy	-0.0647	0.0494	0.8960	SSDxoy	-0.0612	0.0448	0.9090
SSDyox	0.7957	0.2641	0.0000	SSDyox	1.7952	0.4402	0.0000
SOmax	-0.0647	0.0494	0.8960	SOmax	-0.0612	0.0448	0.9090
HH1 (x) vs. HH4 (y)				HH1 (x) vs. HH5 (y)			
FSDxoy	0.0139	0.0246	0.0290	FSDxoy	0.0016	0.0067	0.1120
FSDyox	0.4357	0.0691	0.0000	FSDyox	0.4804	0.0697	0.0000
FOmax	0.0139	0.0246	0.0290	FOmax	0.0016	0.0067	0.1120
SSDxoy	-0.0136	0.0474	0.6240	SSDxoy	-0.0747	0.0474	0.9410
SSDyox	2.7932	0.5951	0.0000	SSDyox	2.7185	0.5164	0.0000
SOmax	-0.0136	0.0474	0.6240	SOmax	-0.0747	0.0474	0.9410
HH2 (x) vs. HH3 (y)				HH2 (x) vs. HH4 (y)			
FSDxoy	0.0040	0.0063	0.0560	FSDxoy	-0.0028	0.0027	0.6890
FSDyox	0.1088	0.0326	0.0000	FSDyox	0.2060	0.0399	0.0000
FOmax	0.0040	0.0062	0.0560	FOmax	-0.0028	0.0027	0.6890
SSDxoy	-0.0753	0.0312	0.9920	SSDxoy	-0.1246	0.0355	1.0000
SSDyox	0.3937	0.1288	0.0000	SSDyox	0.7570	0.1382	0.0000
SOmax	-0.0754	0.0306	0.9920	SOmax	-0.1246	0.0355	1.0000
HH2 (x) vs. HH5 (y)				HH3 (x) vs. HH4 (y)			
FSDxoy	0.0000	0.0001	0.0420	FSDxoy	0.0097	0.0152	0.0420
FSDyox	0.2529	0.0352	0.0000	FSDyox	0.1139	0.0383	0.0000
FOmax	0.0000	0.0001	0.0420	FOmax	0.0094	0.0143	0.0420
SSDxoy	-0.1470	0.0349	1.0000	SSDxoy	-0.0085	0.0346	0.6200
SSDyox	0.8186	0.1208	0.0000	SSDyox	0.5275	0.1973	0.0040
SOmax	-0.1470	0.0349	1.0000	SOmax	-0.0095	0.0325	0.6240
HH3 (x) vs. HH5 (y)				HH4 (x) vs. HH5 (y)			
FSDxoy	0.0019	0.0064	0.1450	FSDxoy	0.0162	0.0198	0.1550
FSDyox	0.1421	0.0344	0.0000	FSDyox	0.0821	0.0309	0.0000
FOmax	0.0019	0.0064	0.1450	FOmax	0.0141	0.0158	0.1550
SSDxoy	-0.0351	0.0310	0.8720	SSDxoy	-0.0275	0.0634	0.8010
SSDyox	0.7267	0.1748	0.0000	SSDyox	0.3513	0.2029	0.0150
SOmax	-0.0351	0.0310	0.8720	SOmax	-0.0358	0.0429	0.8160

Glossary of Tests:
FSDxoy First order stochastic dominance of **x** over y
FSDyox First order stochastic dominance of y over x
FOmax First order maximal
SSDxoy Second order stochastic dominance of **x** over y
SSDyox Second order stochastic dominance of y over x
SOmax Second order maximal
Prob Reject the null of no dominance when the statistics are negative