China's Income Distribution and Inequality

Ximing Wu* and Jeffrey M. Perloff**

January 2004

* Department of Economics, University of Guelph; xiwu@uoguelph.ca.

** Department of Agricultural and Resource Economics, University of California, Berkeley; member of the Giannini Foundation; perloff@are.berkeley.edu.

Abstract

We use a new method to estimate China's income distributions using publicly available interval summary statistics from China's largest national household survey. We examine rural, urban, and overall income distributions for each year from 1985-2001. By estimating the entire distributions, we can show how the distributions change directly as well as examine trends in traditional welfare indices such as the Gini. We find that inequality has increased substantially in both rural and urban areas. Using an inter-temporal decomposition of aggregate inequality, we determine that increases in inequality within the rural and urban sectors and the growing gap in rural and urban incomes have been equally responsible for the growth in overall inequality over the last two decades. However, the rural-urban income gap has played an increasingly important role in recent years. In contrast, only the growth of inequality within rural and urban areas is responsible for the increase in inequality in the United States, where the overall inequality is close to that of China. As a robustness check, we show that consumption inequality (which may be a proxy for permanent income inequality) in urban areas also rose considerably.

Table of Contents

Increasing Chinese Income Inequality Due to a Growing Rural-Urban Income Gap

1. Introduction

Using a new technique to estimate income distributions from grouped summary statistics, we show that Chinese income inequality rose substantially from 1985 to 2001 because of increases in inequality within urban and rural areas and the widening rural-urban income gap. We find that China's dramatic economic growth—a five-fold increase in the economy and a four-fold increase in per capita income since the early 1980s—has disproportionately favored the urban areas and the rich. We also show that the rural and urban income distributions have evolved along separate paths, and this divergence has contributed markedly to the rise in the overall level of inequality.

Although a few articles have reported that income inequality in China increased rapidly over the last two decades, none shows by exactly how much inequality rose because of the absence of consistent, reliable income distribution estimates over time (Bramall, 2001). The Chinese government provides Gini indices for only a few, random years using unspecified data sources, income definitions, and methodologies, hence its inequality measures may not be directly comparable over time (see United Nations, *World Income Inequality Database*).

Moreover, the Gini index only reflects some aspects of the underlying income distribution: A large amount of information is lost. Two Lorenz curves with the same Gini value may have different shapes. As we demonstrate below, although the Gini index of the 1999 U.S. income distribution (0.414) is almost identical to that of 2001 China income distribution (0.415), the shapes of the two distributions differ markedly. Thus, welfare implication from comparing Gini coefficients (or other summary statistics) may be ambiguous. Consequently, we report

several summary statistics as well as reliable estimates of the entire income distribution. Throughout our paper, we compare Chinese to U.S. income distributions to illustrate that, though both countries currently have similar Gini indexes, the reasons these countries are experiencing growing inequality differ.

This paper makes four contributions. First, we use the new method introduced in Wu and Perloff (2003) to estimate flexible income distribution functions when summary statistics are only available by intervals rather than for the entire distribution. Using the income summary statistics based on China's annual national household survey, we estimate rural, urban and overall income distributions for each year from 1985 through 2001. Based on these estimated income distributions, we provide the first intertemporally-comparable series of income inequality estimates of China based on a single consistent data source, methodology, and set of definitions.

Second, we show how the rural, urban, and overall Chinese income distributions evolved over time, and not merely how an arbitrarily chosen summary statistic, such as the Gini, changed. We show that the rural and urban income distributions evolved along different paths. We employ a simple new measure of the overlap between two distributions, which is the area under both density functions: the intersection.

Third, we decompose China's total inequality between rural and urban sectors to explore the distributional impacts of income growth, rural-urban income gap, and migration over time. We show that the rising inequality within both rural and urban areas, the widened rural-urban income gap, and the shift of populations between these two areas were responsible for the rise in aggregate inequality. We show that the widening rural-urban income gap played a major role in China unlike in the United States even though both countries have roughly equal levels of overall income inequality. Migration from rural to urban areas has little effects on the aggregate

inequality in both countries for different reasons. U.S. migration does not affect inequality within either sector or between the two sectors; while Chinese migration affects both within and between inequality significantly, but these effects are offsetting.

Fourth, as a robustness check, we examine the consumption inequality for urban areas. Consumption is a relatively reliable proxy for permanent income. As such, it provides an alternative to the limited income data. We find that the consumption inequality is also rising rapidly in China.

Section 2 discusses possible causes for the increase in China's overall inequality. The following section describes the available data. The fourth section presents our method to estimate maximum entropy densities using grouped data. The fifth section estimates China's income distributions and inequality for 1985-2001. The sixth section shows the relationship between total inequality and rural and urban inequality. The seventh section presents measures of consumption inequality for urban areas as a proxy for permanent income inequality. The last section summarizes our results and presents conclusions.

2. Causes of Increased Inequality

The existing literature (Khan and Riskin 1998, Gustafsson and Li 1999, Yang 1999, Li 2000, and Meng 2003) argues that income inequality has increased markedly in China over the last couple of decades. Khan and Riskin (1998) and Li (2000) also provide limited evidence that China's rural and urban income inequality differ and are growing at different rates.

We will present evidence that the increase in China's overall inequality is due to increases in *within inequality*, the inequality within the rural sector and within the urban sector, and *between inequality*, the inequality due to differences in the average income level between the rural and urban sectors, as well as population shifts between the sectors. Our explanation is a

generalization of two popular explanations—the Kuznets curve hypothesis and the structural hypothesis—which have contrasting implications about future inequality.

Kuznets (1955) stressed the role of between inequality in explaining the evolution of total inequality over time. He hypothesized that, if between inequality is greater than within inequality in each sector, then overall inequality will initially rise as people move from the low-income (rural) sector to the high-income (urban) sector. Later, inequality will fall, as most of the population settles in the high-income, urban sector. The resulting inverted U-shape relationship between inequality and the income level is called a Kuznets curve.[1] If this hypothesis is true, the increase in inequality in developing countries during the course of urbanization may be a transitory process, and inequality will decline at the conclusion of the urbanization process.

A similar explanation starts from the same premise that the rural-urban income gap is the driving force for increased overall inequality, but holds that the adjustments described by Kuznets will not occur due to the secular demographic and institutional structure of China. According to this explanation, China's population has been divided into separate rural and urban economies. To a limited degree, migrants from rural areas may seek jobs in urban areas but China's strict residence registration system usually prevents them from obtaining urban residence status (and hence access to welfare benefits and subsidies enjoyed by urban residents and higher paying jobs). For example, Yang (1999) uses a static "within and between" analysis of household survey data from two provinces for 1986, 1992, and 1994 to argue that increases in rural–urban income differentials is the major cause of rising overall aggregate inequality in

[1] Many studies have estimated the Kuznets curve using cross-country comparisons. Recently this literature has been criticized for failing to account for country-specific effects and for using data that are not comparable. Analyses using panel data from a single country suggest that there is no intrinsic tradeoff between long-run aggregate economic growth and overall equality. See Bruno et al. (1996) for a review of this literature.

China.[2] He suggests that urban-biased policies and institutions are responsible for the long-term rural–urban divide and the recent increase in disparity. If barriers to migration remain, then inequality is unlikely to diminish in the future.

Thus, both of these hypotheses emphasize the rural-urban gap as the primary cause of increasing aggregate inequality. This factor is certainly part of the explanation for growing inequality. However, the complete story is more complex. We will present evidence that, over the last two decades, the increase in both within and between inequality contributed substantially to increased aggregate inequality and that population shifts also affect the increase in overall inequality. In particular, we show that if one takes into account migration, changes in within and between inequality were equally responsible for the increase in overall inequality (in contrast to the traditional static analysis which concludes that between inequality was largely responsible).

Income inequality does not have a clear secular trend. Chang (2002) argues that "… a cure for this problem is to accelerate urbanization in the short run and to promote the growth of the urban sector in the long run. Yet, these policies in the short run may further widen the measured income gap." However, the urban sector may not be able to absorb the large rural surplus workers (150 million according to Chang, 2002) and China's residence registration system may restraint migration into urban areas. Therefore it is likely that China will maintain a high level of income inequality for an extended period.

3. Data

We rely on the largest, most representative survey of Chinese households. The State Statistics Bureau of China (SSB) conducts large-scale annual household surveys in rural and

[2] Because Yang's analysis is restricted to only two provinces for a shorter time period, his results are not directly comparable to our results.

urban areas. The surveys cover all 30 provinces. They usually include 30,000 to 40,000 households in urban areas and 60,000 to 70,000 in rural areas. The SSB uses a two-tier stratified sampling scheme to draw a representative random sample of the population. Each household remains in the survey for three consecutive years. Each year, one-third of the households rotate out of the sample and are replaced by incoming households. Households are required to keep a record of their income and expenditure.

Because we do not have access to the underlying individual data from the SSB survey for all regions and all years, we estimate the Chinese rural and urban income distributions using publicly available summary statistics. Unfortunately, the SSB does not provide summary statistics for the entire sample, but only for various income intervals. These interval summary statistics are published for urban and rural areas in the *Chinese Statistics Yearbook* ("Yearbook" henceforth). The Yearbook defines the family income as annual family disposable income. Our sample covers 1985 through 2001, a period for which the Yearbooks provide consistent data over time.

The Yearbooks summarize the income distributions differently for rural and urban areas. Rural income distribution is divided into a fixed number of intervals. The limits for these income intervals and the share of families within each interval are reported, as is the average income of the entire distribution, but not the conditional mean of each interval. The Yearbooks report 12 rural income intervals for 1985–1994, 11 for 1996, and 20 for 1995 and 1997–2001. For urban areas, the Yearbooks report the conditional mean of the 0-5th, 5-10th, 10-20th, 20-40th, 40-60th, 60-80th, 80-90th, and 90-100th percentiles of the income distribution, but not the limits of these income intervals. We use these publicly available grouped data to estimate the underlying

distributions and draw inequality inferences from estimated income distributions. Both rural and urban income are deflated by the corresponding Consumer Price Index (CPI) from the Yearbook.

4. Maximum Entropy Density Estimation with Grouped Data

Many earlier studies (e.g., Gastwirth and Glauberman 1976, Kakwani and Podder 1976, and Chen et al. 1991) estimated inequality and poverty using grouped data. These papers concentrated on estimating the Lorenz curve and its associated inequality indices. In contrast we use the method developed in Wu and Perloff (2003) that generalizes the traditional maximum entropy density method to estimate a very general income density function using grouped data. ; By so doing, in addition to determining the Lorenz curve and various welfare indices, we can examine how the shape of the entire income distribution and how it changes over time.

The principle of maximum entropy (Jaynes, 1957) is a general method to assign values to probability distributions on the basis of partial information. This principle states that one should choose the probability distribution, consistent with given constraints, that maximizes Shannon's entropy. Traditionally, this maxent density can be obtained by maximizing Shannon's information entropy

$$W = -\int p(x) \log p(x)\, dx$$

subject to K known moment conditions for the entire range of the distribution

$$\int p(x)\, dx = 1,$$
$$\int g_i(x)\, p(x)\, dx = \mu_i, \quad i = 1, 2, ..., K.$$

We can solve this optimization problem using Lagrange's method, which leads to a unique global maximum entropy (Zellner and Highfield, 1988; Ormoneit and White, 1999; and Wu, 2003). The solution takes the form

$$p(x) = \exp\left(-\lambda_0 - \sum_{i=1}^{K} \lambda_i g_i(x)\right),$$

where λ_i is the Lagrange multiplier for the i^{th} moment constraint. This maximum entropy method is equivalent to a maximum likelihood approach where the likelihood function is defined over the exponential distribution and therefore consistent and efficient. See Golan and Judge (1996) for a discussion of how these two approaches are dual.

All the best-known distributions can be described as maxent densities subject to simple moment constraints, which we will call "characterizing moments" henceforth. These characterizing moments are sufficient statistics for exponential families; the entire distribution can be summarized by the characterizing moments.

When only grouped summary statistics are reported, we can estimate the maxent density by incorporating the grouped information as partial moments. Suppose that, for a certain distribution, we only know the grouped summary statistics of M intervals, with interval limits $[l_0, l_1, \ldots, l_M]$, and J conditional moments of each interval

$$\begin{bmatrix} \nu_{1,1} & \nu_{2,1} & \cdots & \nu_{M,1} \\ \nu_{1,2} & \nu_{2,2} & \cdots & \nu_{M,2} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{1,J} & \nu_{2,J} & \cdots & \nu_{M,J} \end{bmatrix} \tag{1}$$

where $\nu_{m,1}$ is the share of the m^{th} interval, and $\sum_{m=1}^{M} \nu_{m,1} = 1$. We define the j^{th} partial moment of a distribution $p(x)$ over the m^{th} interval as

$$\nu_{m,j} = \int_{l_{m-1}}^{l_m} f_j(x)\, p(x)\, dx,\ m = 1,\ldots,M \text{ and } j = 1,\ldots,J.$$

Given the underlying density function is $p(x) = \exp\left(-\lambda_0 - \sum_{i=1}^{K} \lambda_i g_i(x)\right)$, we calculate $p(x)$ using the partial moment conditions.[3] Substituting $p(x)$ into the partial moment conditions, we obtain a system of $(M \times J)$ equations, one for each entry of matrix (1). We can solve for the Lagrange multipliers by iteratively updating

$$\lambda^{(1)} = \lambda^{(0)} + \left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{b},$$

with $b_{m,j} = \nu_{m,j} - \int_{l_{m-1}}^{l_m} f_j(x)\, p(x)\, dx$. The $(M \times J)$ by J matrix $\mathbf{G}$ consists of M submatrices $\mathbf{G}^{(m)}$ $(J \times J)$ stacked on top of one another, where

$$G_{ij}^{(m)} = \int_{l_{m-1}}^{l_m} g_j(x)\, g_k(x)\, p(x)\, dx,\ 1 \le i, j \le J.$$

When the interval limits are unknown, the estimation procedure is more complicated because we do not know over which ranges the conditional means should be evaluated. For example in the Yearbooks, unlike rural areas, only the share and conditional mean of each urban income interval are reported. The moment constraints take the form

$$\begin{aligned} \nu_{m,1} &= \int_{l_{m-1}(\,)}^{l_m(\,)} p(x)\, dx,\ m = 1, \ldots, M, \\ \nu_{m,2} &= \int_{l_{m-1}(\,)}^{l_m(\,)} xp(x)\, dx, m = 1, \ldots, M, \end{aligned} \tag{2}$$

where the interval limits $l_m(\,)$'s are functions of the unknown density function, $p(x)$. Wu and Perloff (2003) show how to estimate the location of these limits using a Quasi-Newton's method, jointly with the density function.

[3] In general, the functional form $p(x)$ is unknown. Wu and Perloff (2003) discuss choosing a model using bootstrapped Kullback-Leibler Information Criterion.

Because we do not have individual Chinese data corresponding to the reported grouped information, we cannot directly examine the effectiveness of the proposed method using Chinese data. Nonetheless, we demonstrate the effectiveness of the proposed method using raw income data from the 2000 U.S. Current Population Survey (CPS) March Supplement: See the Appendix and Wu and Perloff (2003). Using the sequential updating method of model selection described in Wu and Perloff (2003), we find that the specification $p(x) = \exp\left(-\sum_{i=0}^{4} \lambda_i \log(1+x)^i\right)$ gives the best overall fit according to the bootstrapped Kullback-Leibler Information Criterion. This method works extremely well for the U.S. data: The fit is virtually as close as could be obtained with moment conditions for the entire sample. For example, given the population shares and means for 8 intervals but not the interval limits, the estimated distribution had a Gini of 0.413; whereas the Gini based on individual data is 0.414.

5. Rural and Urban Inequality over Time

Using this method, we estimate the Chinese rural and urban income distributions from publicly available summary statistics. Based on these estimated distributions, we use three approaches to determine whether inequality rose over the last two decades. First, we examine how traditional inequality measures changed over time. Second, we examine growth incidence curves, which trace out the growth rate of each income quantile between two points of time. Third, we compare the estimated distributions directly.

Traditional Measures of Inequality

We start by examining three traditional measures of inequality—the Gini Index, the mean logarithm deviation of income, and comparisons of quantile ranges—for rural and urban areas separately. We use these measures to examine how inequality has changed over time.

From the rural survey, we have 12 intervals for 1985–1994, 20 for 1995, 11 for 1996, and 20 for 1997–2001. For urban areas, we have conditional means of the 0-5th, 5-10th, 10-20th, 20-40th, 40-60th, 60-80th, 80-90th, and 90-100th percentiles of the income distribution, but not the limits of these income intervals.[4] We estimate the rural income distribution subject to the proportion of families in each *known* interval. Because the limits for the income intervals are unknown for urban income, we estimate them jointly with the density function. Again, we find that the specification $p(x) = \exp\left(-\sum_{i=0}^{4} \lambda_i \log(1+x)^i\right)$ gives the best overall fit for both areas according to the bootstrapped Kullback-Leibler Information Criterion.

Table 1 shows how our various inequality measures for each year in our samples. The first two columns of numbers report the Gini index for rural and urban areas based on our estimated distributions for each year. The next two columns show the rural and urban mean logarithm deviations ($MLD = \frac{1}{n}\sum_i \log(\mu / x_i)$, where n is the number of people).[5]

According to both measures, rural areas have greater inequality than urban areas throughout the period. On average, the rural Gini is 1.4 times and the *MLD* is 2.2 times their urban counterparts.

---

[4] Consequently, we have more confidence in our rural income distribution estimate than our urban one because the rural distribution is summarized in more intervals (20 versus 8), spans the entire distribution relatively evenly, and has income limits. More importantly, the top urban interval covers the entire 90-100th decile. If most of the dispersion at the high end of the distribution occurred within the top decile during our sample period, we cannot recover this increase without further information.

[5] The *MLD* belongs to the family of generalized entropy index, $I_a = \sum_i \left[(x_i/\mu)^a - 1\right]/[na(a-1)]$, where $a \geq 0$. A low value of a indicates a high degree of "inequality aversion". One can show that $\lim_{a\to 0} I_a = \frac{1}{n}\sum_i \log(\mu / x_i)$, which is the *MLD*. In this study, we focus on the *MLD* as it gives the simplest formula for the intertemporal decomposition of inequality (see Section 5).

The correlation between the Gini and the *MLD* is 0.76 for rural areas and 0.73 for rural areas. Both inequality measures for rural and urban areas increased steadily over the sample period. The rural Gini increased by 26% from 0.272 to 0.343. One reason we are confident that the Gini is capturing a real, upward trend is that we compared the calculated Lorenz curves from the estimated densities. For example, the 1985 Lorenz curves of rural and urban distributions lie above those for 2001 everywhere, suggesting that the 1985 distributions Lorenz dominate those for 2001.[6]

The rural *MLD*—which places a relatively large weight on the income at the low end of the distribution—increased by 67.7% from 0.127 to 0.213. Urban inequality rose faster, though it remained below that in rural areas. The urban Gini increased by 40.8% from 0.191 to 0.269, and the *MLD* nearly doubled from 0.060 to 0.119.

Another traditional approach to assess the changes in inequality is to compare quantile ranges. Because of the interval summary statistics nature of our data, the information loss for quantile estimates due to grouping may be less than that of inequality index of the entire range, which suffers from the aggregating over the top and bottom quantiles. The last four columns of Table 1 show the estimated 90/50 and 50/10 quantile ratios. If $Q(p)$ is the p^{th} percentile, then the 90/50 quantile ratio is $Q(90)/Q(50)$. The 90/50 ratio reflects the relative shares of a wealthy group to the average group. Similarly, the 50/10 quantile ratio shows the relative shares of the average to a poor group. For rural and urban areas, both measures increased by between 20 and 25% during the sample period. Although not shown in the table, the 90/10 ratio increased by around 50%. The similarity in changes of these quantile ratios suggests that the different

[6] Given the estimated density f and sample average μ, the Lorenz curve is obtained numerically as $L(p)=\frac{1}{\mu}\int_0^{F^{-1}(p)} xf(x)\,dx$, where F^{-1} is the inverted distribution function.

inequality increase rate, as measured by Gini and *MLD*, is likely due to the difference in evolutions of the upper and lower tails of the distributions.

Given how China records rural migrants to urban areas, studies based on any Chinese data set measure rural and urban inequality differently than they would in other countries. As migrants from rural who work urban areas usually cannot obtain urban residence status, they are excluded from urban household surveys. Because migrants can only obtain jobs that pay less than those of other urban workers and because the number of migrants grew considerably during the sample period, urban inequality measures are lower than if migrants were counted as urban residents.[7] On the other hand, if migrants earn relatively high incomes by rural standards, including them in the rural household surveys raises rural income inequality. Moreover, Schultz (2003) notes that restrictions on permanent migration reduce the returns that rural youth can expect to realize through profitably moving to a higher wage labor market. Consequently, the household registration system increases the gap in investments in education between rural and urban families and the rural-urban gap in the long run.

Comparison with the Literature

We can compare our estimates to those from four previous studies. As these other studies only report the Gini for a few years, Table 2 compares the rural and urban Gini indexes for only those years.

Li (2000) reports rural and urban Gini index based on SSB micro data for 1988 and 1995. Our estimates of the rural Gini of 0.300 in 1988 and 0.338 in 1995 are close to Li's (2000) estimates based on SSB data of 0.301 and 0.332. Our estimates of the urban Gini of 0.201 in

[7] During the sample period, the share of the rural population fell from 76% to 62%. The number of migrant workers is estimated to be around 80 million in the mid-1990s. See Bramall (2001) and references therein.

1988 and 0.221 in 1995 are not quite as close to Li's estimates, 0.23 and 0.28. As we discussed above, underestimates of urban inequality may be the result of lost information from excessive grouping and top coding as the urban income distribution is summarized by only 8 groups and the highest interval covers the top decile.

Because the SSB household survey data are not publicly available, the other three studies—Khan and Riskin (1998), Gustafsson and Li (1999), and Meng (2003)—use data from smaller, less representative surveys conducted by the Economics Institute of the Chinese Academy of Social Sciences (CASS) in 1988 and 1995.[8] The CASS uses a broader definition of income than does the SSB. Although three of these studies use the CASS data, their estimates of the Gini differ, because they make different assumptions about the underlying data (Bramall, 2001).

Khan and Riskin (1998) report higher rural inequality measures based on CASS data than either we or Li (2000) do based on SSB data. All three CASS studies estimate the 1988 urban Gini at 0.23 (above our estimate of 0.20), but their estimates of the 1995 value range from 0.28 to 0.33 (all higher than our 0.22). Thus, our urban estimates are lower than those of all four previous studies. The lower value of our estimates may be due to difference in the underlying data sources, definitions of income, or methodology.

Nonetheless, all studies report that rural and urban inequality increased from 1988 to 1995. In addition, Meng (2003) reports that the urban Gini increased from 0.282 in 1995 to 0.313 for 1999 based on a CASS survey covering six provinces.

[8] Unlike the SSB survey that covers all 30 provinces, the CASS survey covered 28 provinces for rural areas and 19 provinces for urban areas in 1988, and 10 provinces for rural areas and 11 provinces for urban areas in 1995.

Distributional Impact of Income Growth

Although these inequality measures indicate that inequality increased significantly during the sample period, these indices do not fully describe how the distribution changed. The main problem with using only inequality indices is that the welfare implications depend on the choice of the index because indices differ in the weight they place on various portions of the income distribution (Atkinson, 1970). We need to know how the entire distribution shifted to determine the full social welfare effects.

During the sample period, the average real income more than doubled in rural areas and increased by 169% in urban areas. To examine how this rapid income growth affects the distributions over time, we use the ratio of income corresponding to the same percentile of two distributions. Following Gastwirth (1971) and Ravallion and Chen (2002), we invert the CDF at the p^{th} quantile to obtain the corresponding income

$$Q(p) = F^{-1} = L'(p)\mu, \quad 0 \le p \le 1,$$

where F is the cumulative distribution function, $L'(p)$ is the slope of the Lorenz curve at the p^{th} quantile, and μ is the overall average income. The growth incidence curve (GIC) between time t–1 and time t is

$$\mathrm{GIC}(p) = \frac{L_t'(p)\mu_t}{L_{t-1}'(p)\mu_{t-1}} - 1 = \frac{Q_t(p)}{Q_{t-1}(p)} - 1, \quad 0 \le p \le 1.$$

It traces out the growth rate of each income quantile between t-1 and t.

If the Lorenz curves do not change during this period, the GIC is equal to the average growth rate (μ_t/μ_{t-1}) everywhere so that growth is neutral. The growth is said to be *pro-rich* if the GIC is upward sloping and *pro-poor* if the GIC slopes down. If the GIC is everywhere above

zero, then the distribution of time t Lorenz dominates that of time t–1. In other words, the Lorenz curve of the second distribution lies strictly above that of the first one.

Figure 2 plots the GIC between 1985 and 2001, divided by the number of years in between, for both areas. Compared with urban areas, the incomes of the poorest rural group grew slightly slower, but that of the middle rural group grew slightly faster. The ratio is everywhere above zero, so all income groups benefited in absolute terms. The rich grew proportionately richer during this period, as the curves are almost everywhere increasing.[9] For rural areas, the annual growth rate for the poorest group is about 3%, while that for the richest is nearly 9%.

We note that the annual average and median growth rates for both sectors are about 7.4%. Thus, the estimated growth rates based on summary statistics of micro household surveys agree with the government number of per capita GDP growth during this period, which is about 7% to 8%.

Examine Distributions Directly

Although it provides a straightforward way to examine changes in the distributions over time, the GIC only reflects certain aspects of the evolutionary process. For example, the GIC analysis does not show how the general shape of the income distribution changed over time. Is the increased inequality as measured by the Gini or *MLD* caused by a rightward shift of the mode, a thickened tail, or some other more complex change? Does the distribution become bi-modal due to "hollowing out" of the middle class? For further insight into this process, we examine the

[9] The bent-down section at the high end of the urban distribution may be due to top-coding in survey of the highest income group and under-reporting of their income by the rich. Both of these effects are presumably more important in the richer urban area than in the rural area.

shapes of our estimates of the flexible density function, which allows for multi-modal distributions.

Figures 3 shows how the rural distribution changed between 1985 and 2001, and Figure 4 shows the shift in the urban distribution. Throughout the sample period, each distribution has a single mode. However, dispersion increased considerably over time, largely because the right tails grew longer. Moreover, the income distributions gradually but persistently moved to the right (and correspondingly, the weight at the mode decreased), reflecting a general increase in incomes.

These rightward shifts in the distributions are more clearly seen by comparing distributions for pairs of years. The left panel of Figure 5 shows that the 2001 rural income distribution is much more dispersed than the 1985 distribution. The distribution mode rose 68% from 292 Yuan in 1985 to 490 (in 1985) Yuan in 2001. Despite the rightward shift of the mode, the skewness increased from 1.28 to 1.39. The height of the distribution at the mode in 2001 is only about 40% of the 1985 peak, which caused kurtosis to fall from 4.95 to 4.86.

The level and the dispersion of the urban income (right panel of Figure 5) rose more rapidly than in rural areas (left panel). Moreover, the fraction of households with very low levels of income fell substantially. The mode of the urban distribution increased by 140% from 681 Yuan in 1985 to 1,634 in (1985) Yuan in 2001, while the density of the mode in 2001 fell to 25% of that in 1985. The distribution became more symmetric—skewness decreased from 1.82 to 1.47—reflecting a relative decrease in the share of poor and increase in the share of wealthy people. The kurtosis fell from 8.28 to 6.05, reflecting the substantial flattening of the peak. Compared with the rural distribution, the share of people with low absolute income (the height of

the left tail) was much smaller, which helps to explain why our inequality estimates are lower in urban areas, especially for the *MLD*, which heavily weights the income of the poor.

By how much did the distributions shift? We can assess the overall distance or closeness between two distributions directly. We propose a simple new measure of the overlap between two distributions, the intersection, which is the area under both density functions. This statistic for two density functions $p(x)$ and $q(x)$ on the real line or its subsets is defined as

$$\Omega = \int \min\left[p(x), q(x)\right] dx,$$

whose value is equal to area *B* in Figure 5.[10] It is restricted to lie within $[0,1]$. If $\Omega = 0$, then $p(x)$ and $q(x)$ are disjoint. If $\Omega = 1$, then $p(x)$ and $q(x)$ coincide.

For the 1985 and 2001 distributions, Ω is higher for rural areas, 0.54, than for urban areas, 0.24, because the urban distribution shifted right by considerably more did the rural one.

6. Decomposition of Aggregate Inequality

What effect do these unequal shifts in the rural and urban distributions have on overall inequality? To answer this question, we decompose the total Chinese inequality between rural and urban areas. Our results suggest that increased inequality within either sector and between sectors contributed to the increase of total inequality.

Aggregate Distribution and Inequality

We compute China's aggregate income distribution as a population-weighted mixture of the rural and urban distributions. We use the resulting distribution to calculate the inequality

[10] Compared with another commonly used distance measure such as the Kullback-Leibler distance, $\int p(x) \ln\left[p(x)/q(x)\right] dx$, our measure has three advantages. First, it has an intuitive graphic interpretation as the overlapping areas of two distributions. Second, and more important, it is symmetric in the sense that Ω is invariant to the order of $p(x)$ and $q(x)$: $\Omega_{p,q} = \Omega_{q,p}$. Third, this index can be used to compare directly more than two distributions.

indices of the aggregate distribution. Denoting rural and urban income distribution as $p_r(x)$ and $p_u(x)$ respectively, we obtain the aggregate distribution by taking their weighted sum:

$$p(x) = s_r p_r(x) + s_u p_u(x), \tag{3}$$

where s_r and s_u is the share of rural and urban population. During the sample period, the share of urban population increases steadily from 24% to 38%.

Figure 6 illustrates the relationship of the aggregate distribution (solid) to the rescaled rural (dot) and urban (dash-dot) distributions for 1985 and 2001. The rural and urban densities are rescaled by their corresponding population weights so that the areas below these two curves sum to one. By comparing the 1985 and 2001 figures, we see that the overall shape of the aggregate distribution was relatively unchanged over the sample period, but the right tail became thicker (note that the scale of the two diagrams differ). The left tail of the 1981 aggregate density is almost completely coincident with the rural density (urban dwellers are not that poor) while both the rural and urban densities span the right tail. In 2001, the urban density is almost entirely responsible for the right tail of the aggregate density.

Table 3 reports the Gini index (second column) and the *MLD* (third column), which were calculated from the estimated aggregate $p(x)$. Over the sample period, the Gini index increased 34% from 0.310 to 0.415, and the *MLD* nearly doubled from 0.164 to 0.317. The overall inequality is much higher than either rural or urban inequality because of the substantial rural–urban income gap. As shown by Equation (3) and Figure 6, the increased aggregate inequality was due to changes in the rural or urban distributions, their interaction (the degree to which the two distributions overlap), and the population weights.

Decomposition of Aggregate Inequality

If an inequality index can be decomposed into within sector inequality and between sector inequality without an interaction term for the overlap of sectors, we can derive the aggregate inequality index from the indexes for the subgroups of the population. The most commonly used inequality index, the Gini, is not decomposable in this sense, so generally we cannot calculate the aggregate Gini index from the Gini indices of its subgroups. However, the *MLD* is decomposable, so we can use the rural and urban *MLD*'s to derive the aggregate *MLD*, and we can show which factors contributed to the growth of the aggregate *MLD* over time.

The decomposition formula for the *MLD* index is

$$\begin{aligned} MLD &= \sum_k s_k MLD_k + \sum_k s_k \log\left(\frac{\mu}{\mu_k}\right) \\ &= MLD_W + MLD_b, \end{aligned} \tag{4}$$

where MLD_k is the inequality for the k^{th} subgroup (here, k = rural or urban), μ_k is the mean income of the k^{th} subgroup, and s_k is the population share of the k^{th} subgroup. The first term, MLD_w, is the *within inequality*: the inequality within the rural or urban sector. The second term, MLD_b, is the *between inequality*: the inequality due to differences in the average income level between rural and urban areas.[11]

Both within inequality and between inequality measures increased considerably during the sample period (last two columns of Table 3). Between inequality increased by more in both relative and absolute terms than within inequality. Between inequality increased by 163% from

[11] For example, suppose x1 = [1, 2] and x2 = [3 , 4, 5]. Using the formula, $MLD = \frac{1}{n}\sum_i \log\left(\mu / x_i\right)$, we calculate MLD_1 = 0.5[log(1.5/1) + log(1.5/2)] = 0.06 and similarly MLD_2 = 0.02. Using Equation (4), $MLD_w = 0.4MLD_1 + 0.6MLD_2 = 0.04$. We can calculate MLD_b = MLD(1.5, 1.5, 4, 4, 4) = 0.1 because, if we give every member of a group its group average, then the inequality of the entire population is the between inequality. Finally, $MLD_w + MLD_b = 0.14 = MLD(1, 2, 3, 4, 5)$.

0.053 to 0.139, while within inequality increased by only 61% from 0.111 to 0.178. As a result of both of these increases, total *MLD* inequality more than doubled.

In Table 4, we show inequality increased over the entire period and in three subperiods. To avoid year-to-year fluctuations, we combine the distribution estimates of each two adjacent years and examine the changes for the entire period and three five-year sub-periods, 1985-86 through 1990-91, 1990-91 through 1995-96, and 1995-96 through 2000-01. The first three columns of Table 4 report the *annual* change in aggregate inequality for the entire period and three subperiods.[12] During the sample period, the overall *MLD* inequality increased from 0.16 to 0.32. Although the average annual increase over the entire period was 0.01, the annual rate of increase rose over time, so that the average increase in the third subperiod was more than doubled that in the first subperiod.

In the first subperiod, the contribution of changes in within (0.36) and between (0.32) inequality to the change in aggregate inequality are roughly equal. However, during the second and third subperiods, the between inequality's contribution increased relative to the within inequality. For the entire period, the increase in between inequality accounts for about 58% (≈0.56/0.98) of the total increase.

Equation (4) shows that three factors contribute to total inequality: the inequality within each subgroup (MLD_k), the relative average income of each subgroup (μ_k/μ), and the population shares of each subgroup (s_k). During the sample period, the share of rural population fell from 76% to 62%. However, the simple "within and between" analysis does not separate the impact of changes in population structure from that of changes in the distribution of each sector.

[12] For example, the change for the first subperiod is calculated as $(MLD_{1990} + MLD_{1991})/2 - (MLD_{1985} + MLD_{1986})/2$ divided by 5, the number of years in the subperiod.

Following Mookherjee and Shorrocks (1982), we differentiate the static "within and between" decomposition to examine the effects of each component directly. Applying the difference operator to both sides of Equation (4), we obtain

$$\begin{aligned}\Delta MLD &= MLD_t - MLD_{t-1} \\ &= \Delta\left(\sum_k s_k MLD_k\right) + \Delta\left(\sum_k s_k \log\left(\frac{\mu}{\mu_k}\right)\right) \\ &\cong \sum_k \overline{s}_k \Delta MLD_k + \sum_k \Delta s_k \, \overline{MLD}_k + \sum_k \left(\overline{\eta}_k - \overline{s}_k\right)\Delta \log\left(\mu_k\right) + \sum_k \Delta s_k \left(\overline{\lambda}_k - \overline{\log\left(\lambda_k\right)}\right) \\ &= \underbrace{\theta_w + \theta_{sw}}_{\Delta MLD_w} + \underbrace{\theta_b + \theta_{sb}}_{\Delta MLD_b},\end{aligned} \tag{5}$$

where $\lambda_k = \mu_k/\mu$, $\eta_k = s_k\lambda_k$, and a horizontal bar over a variable indicates that two periods are averaged. We further decompose the contribution from within inequality or between inequality into two components: a pure within or between effect and an effect caused by a change in population shares. The last line of Equation (6) shows that the change in *MLD* is the sum of four effects: θ_w, the effect from changes in within inequality should the population structure remain constant; θ_{sw}, the effect of changes in population structure on within inequality; θ_b, the effect from changes in between inequality (the average income of each group) should the population structure remain constant; and θ_{sb}, the effect from changes in population structure on between inequality. Therefore, by explicitly accounting for the effects of changes in population structure, we are able to separate the contribution of each factor to the aggregate inequality.

We calculate the intertemporal decomposition between each pair of adjacent years (that is, we examine the aggregated income distribution of each two adjacent years to avoid the effects of year-to-year fluctuations on the analysis.) The last four columns of the top panel in Table 4 report the *annual* change in aggregate inequality and each term in Equation (5) for the entire period and three sub-periods. The results suggest that the relative contribution of within inequality ignoring population shifts, θ_w, is larger than the static measure of the change of within

inequality, $\Delta MLD_w = \theta_w + \theta_{sw}$, which includes the effects of the changing population (θ_{sw}). That is, migration from higher-inequality rural areas to lower-inequality urban areas reduces the effect of rising within inequality. On average for the entire period, migration partially offsets the effect of increased within inequality by 17% (= -0.08/0.48).

In contrast, the contribution of between inequality—the rural-urban income gap—is smaller when we account for change in population shares. Because of the widening rural-urban income gap, migration enhances the effect of increased between inequality by 19% (= 0.09/0.47) on average.

The effects of migration on the within and between inequality are nearly offsetting ($\theta_{sw} + \theta_{sb} = 0.01$). Overall, the static "within and between" decomposition underestimates the contribution of increased within inequality because it fails to take into account the influence of change in population structure. For the entire period, the change in within and between inequality each contributes about 50% to the increase in total inequality, compared to 42% and 58% in the simple "within and between" decomposition.

The pattern varies over time. Initially within inequality played a larger role; but in recent years, between inequality contributed more to overall inequality change. After controlling for the effects of migration, we find that changes in within inequality were responsible for 63% of the change in total inequality for the late 1980s; the two components were equally important in the early 1990s; and between inequality played a larger role (55%) in the late 1990s. It is in the late 1990s that the most dramatic increase in inequality occurs. The annual increase in aggregate inequality is 0.014 in the *MLD*, compared with 0.0068 and 0.008 for the first two sub-periods.

Comparison with the United States

Comparing the determinants of changes in Chinese rural, urban, and aggregate income distributions to those in the United States may illustrate the difference between a developing and an industrial economy with currently similar levels of income inequality. We conduct the same intertemporal between-within analysis using U.S. data: the March Current Population Survey (CPS) for 1985-2001. We look at the change in inequality for the entire period as well as for three five-year subperiods. The results are reported in the bottom panel of Table 4.

One important effect that is common to both the United States and China is that inequality is increasing rapidly in both rural and urban areas, which drives up overall inequality. However, China's growing rural-urban income gap and increasing migration into urban areas further forces inequality to rise. For the same period, U.S. inequality in both sectors increased considerably and almost all the changes in overall inequality are attributed to these changes in within inequality. In contrast to the pattern in China, the U.S. share of urban population (70%) and the rural-urban income ratio (75%) have remained relatively constant. Considering the relatively small share of rural population and the stable rural-urban income ratio, neither between inequality nor migration has played a significant role in the rise in U.S. overall inequality. With the share of urban population stable for an extended period, Kuznets' the migration/urbanization process appears to have come to a conclusion. However, instead of going down, the overall inequality has been rising steadily due to the increased inequality within each sector.

7. Consumption Inequality

Because we have been relying on highly aggregate income information, we consider an alternative approach in which we examine Chinese consumption inequality as a proxy for permanent income inequality. Consumption data are only available for urban areas, where

consumption information is summarized in the same format as is income distribution by the Yearbooks.

Jorgenson (1998) argues that estimates of welfare indices depend critically on the choice between income and consumption as a measure of household resource. Permanent income may be the preferred indicator of household resource, but it is unobservable. Although measured income is correlated with permanent income, its substantial transitory component is uncorrelated with permanent income. Measured consumption can serve as a proxy for household permanent income, if it is proportional to permanent income. Moreover, it exhibits relatively smaller transitory fluctuation. Therefore, we may be able to make more reliable welfare inferences using consumption rather than income.

According to several studies of inequalities in the OECD countries report, the recent rise in income inequality was not accompanied by a similar increase in consumption inequality. These findings are sometimes cited in response to public concern about rising income inequality.

Regardless of the validity of this argument in OECD countries, it does not apply to China, where the income and consumption inequality measures are highly correlated. Figure 7 compares the estimated Gini index for income and consumption in the left panel and their growth rate in the right panel. Although consumption inequality is lower than income inequality, its growth rate closely parallels that of the income inequality. In contrast, Krueger and Perri (2002) report that, although the U.S. income Gini index rose substantially from 0.31 to 0.41 during the last quarter of the twentieth century, the consumption Gini index rose 2 percentage points from roughly 0.25 to 0.27. During the 1990s when the income inequality increased considerably, the consumption inequality actually declined.

Prior to 1997, the ratio of average expenditure to average income for households within the 0-5th percentiles of the income distribution averaged 1.06. Hence, consumption by households with very low income exceeded their income, probably due to government subsidies for urban residents. However, the consumption–income ratio for the bottom five percentiles fell to 0.96 for 1997–2001, suggesting that the safety net for the poor may not be as effective as it formerly was. The (relative) deterioration of the consumption of those at the low end of the income distribution and the subsequent rapid increase in consumption inequality near the end of the sample during the late 1990s may be partially due to the large number of workers in the state-owned enterprises who were laid off with only nominal unemployment compensations.[13] The state public-transfer system failed to provide them with the much-needed "safety net". China's government transfers as a share of GDP decreased from 0.35% in 1985 to 0.28% in 2001. In contrast, Keane and Prasad (2003) observe that, unlike most other transition countries, Poland experienced very little increase in overall income inequality. The main reason was that, during the earlier years of transition, there was a sharp increase in social transfers, from about 10% of GDP to 20%.

8. Summary

We examine the evolution of China's income distribution and inequality from 1985 through 2001. We estimate China's income distribution using a new maximum entropy density approach that works well when only a limited set of summary statistics by income interval are available. The maximum entropy principle is a general method to assign values to probability distributions on the basis of partial information. We extend this method to grouped data and use

[13] Reportedly, 11.57 million workers were laidoff in 1997 (China Development Report, 1998).

it on summary statistics of income data from annual Chinese household surveys. We are able to confirm that this new method works extremely well on U.S. data.

Using this new technique and data from the most inclusive Chinese survey, we are able to provide the first intertemporally consistent estimates of China's inequality measures. In contrast to, most previous studies of Chinese income inequality used an alternative survey that is only available in a couple of years and that does not cover the entire country.

We find that rural and urban inequality have increased substantially. Urban inequality was lower than rural inequality during the sample period, but it is rising faster. Direct examination of the estimated distributions reveals that both rural and urban income distributions are shifting to the right over time. The overall dispersion increased considerably, due in large part to the growth of the right tail of the distribution and the failure of the share of the very poor to decline significantly. Although most people's incomes rose over time, the rapid income growth favored the richest members of society, who enjoyed a larger increase in income than did the poor.

Rising inequality within rural and urban areas, the widening rural-urban income gap, and shifts of population between urban and rural areas combined to drive up the aggregate inequality substantially. In contrast to previous studies that used static decompositions that attributed the growth in overall inequality largely to increases in the rural-urban gap, our dynamic decomposition shows that the increase in within and between inequality contributed equally to the rise in overall inequality over the last two decades. We do find, however, that the rural-income gap has played an increasingly important role in recent years.

Finally, we find that consumption inequality, arguably a better indicator of economic well-being than China's noisy income information, has also risen substantially during the sample period. Thus, we are even more concerned that inequality is rising rapidly in China.

In short, Chinese rural, urban, and overall income inequality are high (comparable to that in the United States) and rising due to increases in within and between inequality. Currently rural incomes are less equally distributed than urban incomes. However, urban inequality is increasing faster than rural inequality. Should this trend continue, urban inequality will eventually overtake rural inequality. Combined with the increasingly widening rural–urban income gap, this trend could further accelerate the increase in inequality as people move to urban areas. Government restrictions limit migration from rural to urban areas. Even if such migration were permitted, it probably is not possible for the urban economy to accommodate the majority of the gigantic rural population. Thus, in contrast to the prediction of the Kuznets' curve, gaps between rural and urban incomes may persist and cause overall inequality to rise for an extended period.

References

Atkinson, A. B. "On the Measurement of Inequality." *Journal of Economic Theory*, 1970, *2*: 244-63.

Bramall, Chris**.** "The Quality of China's Household Income Surveys." *The China Quarterly*, 2001, *167*: 689-705.

Bruno, M; Ravallion, M and Squire, L**.** "Equity and Growth in Developing Countries: Old and New Perspectives on the Policy Issues." Working Paper, World Bank, 1996.

Chang, Gene H. "The Cause and Cure of China's Widening Income Inequality." *China Economic Review*, 2002, *13*: 335-40.

Chen, Shaohua; Ravallion, Martin and Datt, Gaurav**.** "POVCAL -- a Program for Calculating Poverty Measures from Grouped Data." Memo, World Bank, 1991.

Chen, Shaohua and Wang, Yan**.** "China's Growth and Poverty Reduction: Recent Trends between 1990 and 1999." Working Paper, World Bank, 2001.

Gastwirth, Joseph**.** "A General Definition of the Lorenz Curve." *Econometrica*, 1971, *39*(6): 1037-39.

Gastwirth, Joseph and Glauberman, Marcia**.** "The Interpolation of the Lorenz Curve and Gini Index from Grouped Data." *Econometrica*, 1976, *44*(3): 479-83.

Golan, A.; Judge, G. and Miller, D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: John Wiley and Sons, 1996.

Gustafsson, B., and Li, S**.** "A More Unequal China? Aspects of Inequality in the Distribution of Equivalent Income." Unpublished Manuscript, 1999.

Kakwani, N.C., and Podder, N. "Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations." *Econometrica*, 1976, *44*(1): 137-48.

Jaynes, E. T. "Information Theory and Statistical Mechanics." *Physics Review*, 1957, *106*: 620-30.

Jorgenson, Dale**.** "Did We Lose the War on Poverty?" *Journal of Economic Perspectives*, 1998, *12*(1): 79-96.

Keane, Michael P. and Prasad, Eswar S. "Social Transfers and Inequality During the Polish Transition," In *Inequality and Growth: Theory and Policy Implications,* edited by T. S. Eicher and S. J. Turnovsky. MIT Press, 2003, 121-54.

Khan, Azizur Rahman, and Riskin, Carl**.** "Income Inequality in China: Composition, Distribution and Growth of Household Income, 1988 to 1995." *The China Quarterly*, 1998, *154*: 221-53.

Krueger, Dirk and Perri, Fabrizio**.** "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." Working Paper, NBER, 2002.

Kuznets, Simon**.** *Shares of Upper Income Groups in Income and Savings*. New York: National Bureau of Economic Research, 1953.

Li, S**.** "Changes in Income Inequality in China Transition." Mimeo, Chinese Academy of Social Sciences, 2000.

Meng, X**.** "Economic Restructing and Income Inequality in Urban China." Research Paper, Australian National University, 2003.

Milanovic, B**.** "Poverty, Inequality and Social Policy in Transition Economies." Research Paper, World Bank, 1995.

Mookherjee, Dilip and Shorrocks, Anthony**.** "A Decomposition Analysis of the Trend in U.K. Income Inequality." *Economic Journal*, 1982, *92*(368): 886-902.

Ormoneit, Dirk and White, Halbert**.** "An Efficient Algorithm to Compute Maximum Entropy Densities." *Econometric Reviews*, 1999, *18*(2): 141-67.

Ravallion, Martin, and Chen, Shaohua. "When Reform Is Faster than Statistical Reforming: Measuring and Explaining Income Inequality in Rural China." *Oxford Bulletin of Economics and Statistics*, 1999, *61*(1):33-56.

Ravallion, Martin, and Chen, Shaohua**.** "Measuring Pro-Poor Growth." Research Paper, World Bank, 2002.

Schultz, T. Paul**.** "Human Resources in China: The Birth Quota, Returns to Schooling, and Migration." Working Paper, Yale University, 2003.

Wu, Ximing**.** "Calculation of Maximum Entropy Densities with Application to Income Distribution." *Journal of Econometrics*, 2003, *115*: 347-54.

Wu, Ximing and Perloff, Jeffrey**.** "Maximum Entropy Density Estimation with Grouped Data." Working Paper, 2003.

Yang, Dennis Tao**.** "Urban-Biased Policies and Rising Income Inequality in China." *American Economic Review Papers and Proceedings*, 1999, *89*(2): 306-10.

Zellner, Arnold and Highfield, Richard A. "Calculation of Maximum Entropy Distribution and Approximation of Marginal Posterior Distributions." *Journal of Econometrics*, 1988, *37*: 195-209.

Appendix

Numerical Example of Maximum Entropy Distributions for Grouped Data

We demonstrate the effectiveness of the proposed method using raw income data from the 2000 U.S. Current Population Survey (CPS) March Supplement. The March CPS, a large annual demographic file with 35,297 observations, includes labor market and income information for the previous year, so the data pertain to tax year 1999.

Corresponding to the different ways the income distributions are summarized in the *Chinese Statistical Yearbook*, we run three experiments. To be consistent with the China data, we divide the U.S. income into 12 and 20 intervals respectively. We then estimate the maxent densities $p_1(x)$ based on 12 intervals and $p_2(x)$ based on 20 intervals, using the corresponding interval limits and share of families in each interval. In the third experiment, we calculate the conditional mean of the 0-5th, 5-10th, 10-20th, 20-40th, 40-60th, 60-80th, 80-90th, and 90-100th percentiles of the income distribution. We then estimate the maxent density, $p_3(x)$, subject to the share and conditional mean of each interval, but do not use the interval limits. We find that the specification $p(x) = \exp\left(-\sum_{i=0}^{4} \lambda_i \log(1+x)^i\right)$ produces the best fit according to the BIC. We compare the estimated densities using two standard measures of inequality, the Gini index and the mean logarithm deviation (*MLD*), where we rescale the income by dividing x by \$10,000. The Gini index and *MLD* from both the raw data and the estimated densities are reported in Table A1. The estimates from the fitted densities are close to those obtained from the full sample.

Table A1. Estimated inequality indices

	Full sample	p_1	p_2	p_3
Gini	0.414	0.409	0.418	0.413
MLD	0.338	0.335	0.348	0.333

In the third experiment, because the limits for the income intervals are unknown, we estimate them jointly with the parameters of the density. The results (in tens of thousands of dollars) are reported in Table A2. They are close to the corresponding sample quantiles.

Table A2. Estimated quantiles

Quantile	5^{th}	10^{th}	20^{th}	40^{th}	60^{th}	80^{th}	90^{th}
Sample	0.097	0.146	0.226	0.386	0.580	0.865	1.154
Estimated	0.092	0.147	0.232	0.384	0.566	0.879	1.226

We can also compare the estimated densities directly using graphs. Figure A1 plots the estimated p_1, p_2, and p_3 against the histogram of the full sample. Our estimated maxent densities successfully capture the shape of the empirical distribution.

Table 1. Estimated Inequality Indices for Rural and Urban Areas

Year	*Gini*		*MLD*		*50/10 Ratio*		*90/50 Ratio*	
	Rural	*Urban*	*Rural*	*Urban*	*Rural*	*Urban*	*Rural*	*Urban*
1985	0.272	0.191	0.126	0.060	1.887	1.478	1.900	1.529
1986	0.284	0.189	0.141	0.059	2.011	1.493	1.956	1.515
1987	0.279	0.194	0.135	0.062	1.976	1.488	1.945	1.533
1988	0.300	0.201	0.160	0.064	2.088	1.524	2.004	1.564
1989	0.305	0.198	0.165	0.063	2.113	1.530	2.064	1.572
1990	0.288	0.198	0.145	0.064	2.012	1.533	1.991	1.569
1991	0.315	0.184	0.178	0.054	2.16	1.483	2.08	1.527
1992	0.317	0.200	0.178	0.065	2.128	1.553	2.126	1.58
1993	0.319	0.219	0.178	0.077	2.123	1.605	2.196	1.682
1994	0.300	0.229	0.156	0.085	2.08	1.661	2.123	1.721
1995	0.338	0.221	0.206	0.079	2.301	1.629	2.205	1.683
1996	0.316	0.221	0.154	0.079	2.123	1.629	2.055	1.690
1997	0.322	0.232	0.168	0.087	2.087	1.682	2.105	1.728
1998	0.321	0.239	0.184	0.093	2.219	1.715	2.147	1.755
1999	0.325	0.246	0.188	0.099	2.227	1.746	2.164	1.790
2000	0.339	0.258	0.210	0.109	2.373	1.791	2.245	1.843
2001	0.343	0.269	0.213	0.119	2.367	1.839	2.301	1.887

Table 2. Comparison of Gini Coefficients

		Gini			
		Rural		*Urban*	
Source	*Data Set*	1988	1995	1988	1995
This study	SSB	0.300	0.338	0.201	0.221
Li (2000)	SSB	0.301	0.323	0.23	0.28
Khan and Riskin (1998)	CASS	0.338	0.416	0.233	0.332
Gustafsson and Li (1999)	CASS			0.228	0.276
Meng (2003)	CASS			0.234	0.282

Table 3. Total Inequality and Its Decomposition

		MLD		
Year	*Gini*	*total*	*within*	*between*
1985	0.310	0.164	0.111	0.053
1986	0.311	0.169	0.121	0.048
1987	0.317	0.175	0.117	0.058
1988	0.337	0.201	0.135	0.066
1989	0.342	0.208	0.138	0.070
1990	0.327	0.186	0.124	0.062
1991	0.345	0.215	0.144	0.070
1992	0.361	0.231	0.147	0.084
1993	0.380	0.255	0.150	0.105
1994	0.381	0.252	0.136	0.116
1995	0.382	0.266	0.169	0.096
1996	0.349	0.215	0.131	0.084
1997	0.375	0.258	0.143	0.116
1998	0.378	0.257	0.154	0.103
1999	0.389	0.272	0.157	0.115
2000	0.407	0.305	0.174	0.131
2001	0.415	0.317	0.178	0.139

Table 4. Contribution of each factor to change in total inequality

Year	ΔMLD	ΔMLD_w	ΔMLD_b	θ_w	θ_{sw}	θ_b	θ_{sb}
China (1985—2001)							
1985-86—1990-91	0.68	0.36	0.32	0.41	-0.04	0.25	0.06
1990-91—1995-96	0.80	0.32	0.48	0.39	-0.06	0.40	0.08
1995-96—2000-01	1.41	0.51	0.90	0.64	-0.13	0.77	0.12
1985-86—2000-01	0.96	0.40	0.56	0.48	-0.08	0.47	0.09
U.S. (1985—2001)							
1985-86—1990-91	0.24	0.27	-0.03	0.27	0.01	-0.03	-0.01
1990-91—1995-96	0.90	0.89	0.01	0.87	0.02	0.02	-0.01
1995-96—2000-01	-0.23	-0.23	0.01	-0.24	0.01	0.01	0.01
1985-86—2000-01	0.30	0.31	-0.01	0.30	0.01	-0.01	-0.01

Note: All numbers have been multiplied by 100.

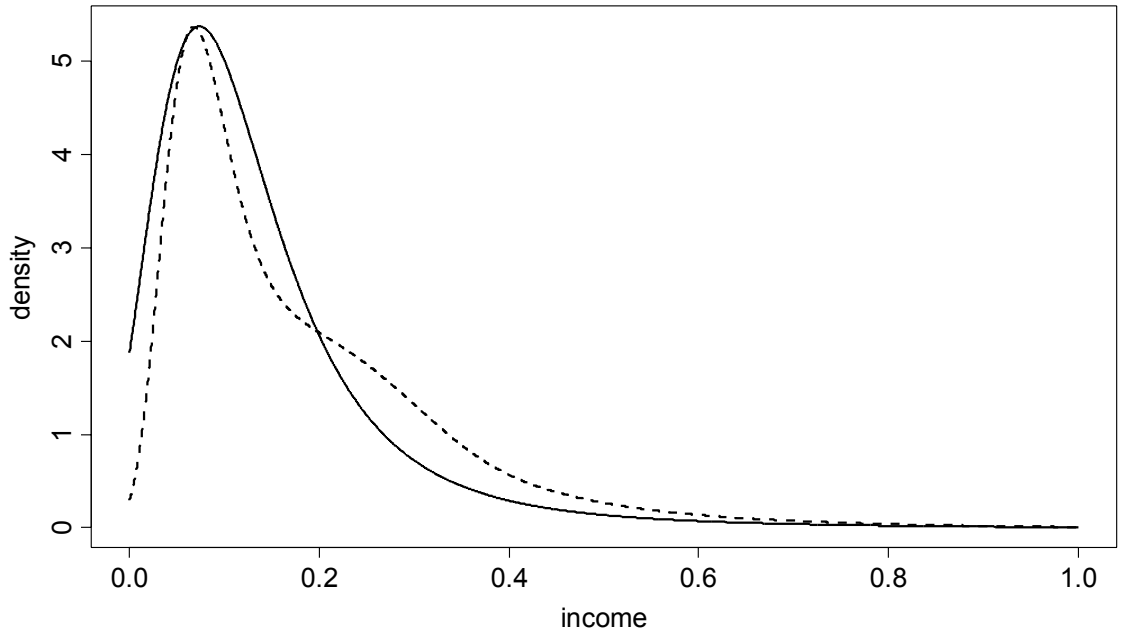

Figure 1. 1999 U.S. income distribution and 2001 China income distribution
(U.S.: solid; China: dashes)
Note: The domains of both distributions have been re-scaled to lie within [0, 1]

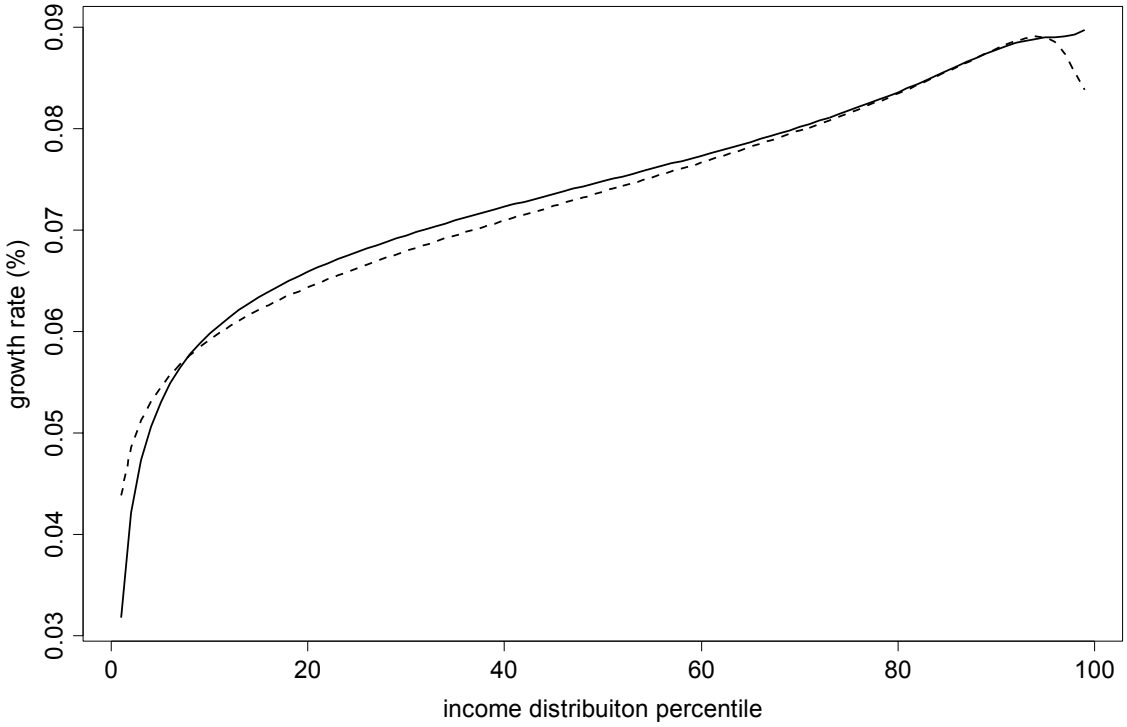

Figure 2. Growth incidence curve, 2001 vs. 1985
(rural: solid; urban: dashes)

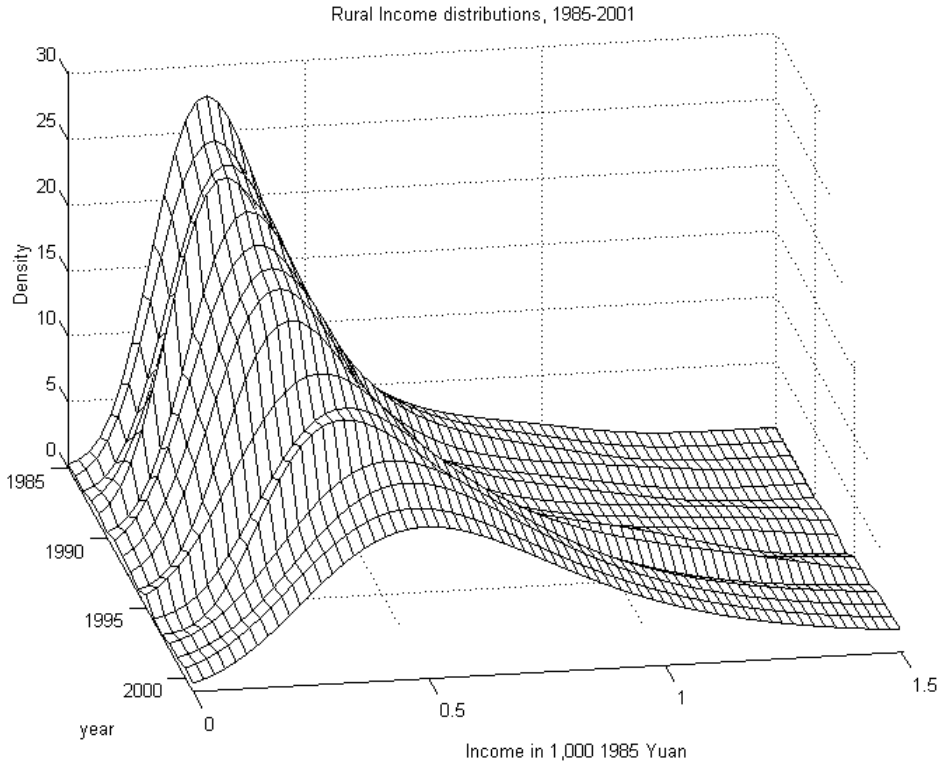

Figure 3. Rural income distributions, 1985-2001

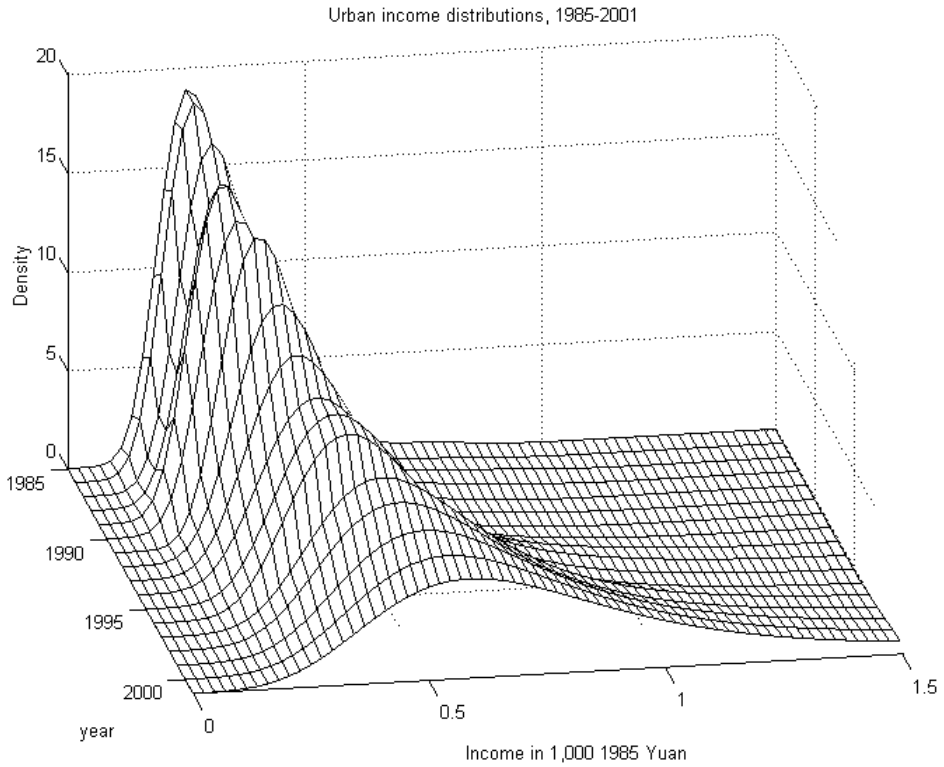

Figure 4: Urban income distributions, 1985-2001

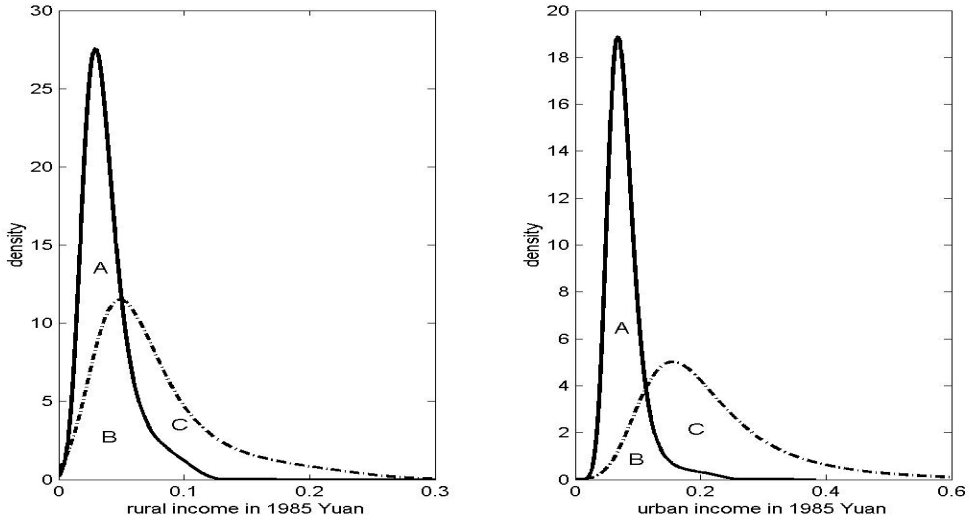

Figure 5. Estimated rural and urban distributions in 1,000 1985 Yuan
(1985: solid: 2001: dashes and dots)

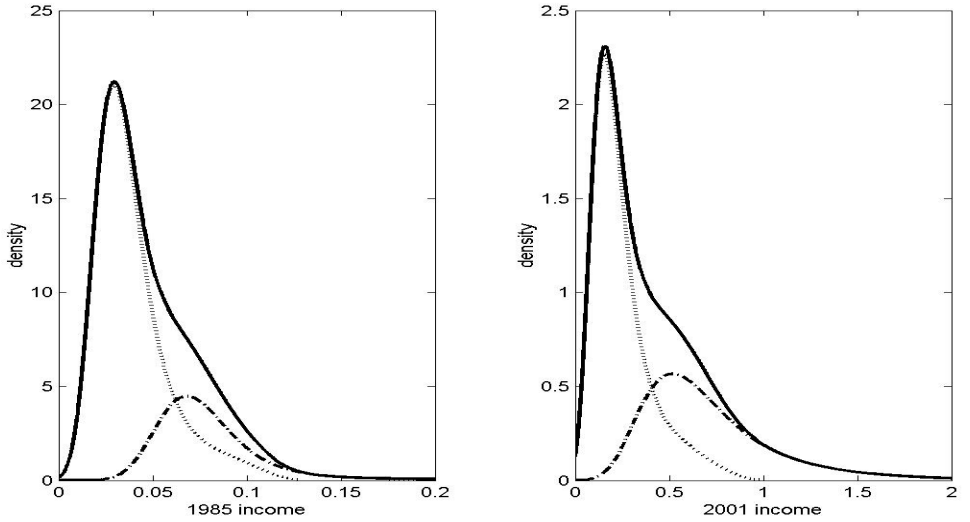

Figure 6. Rural, Urban, and Aggregate distributions for 1985 and 2001 (in 1,000 current Yuan)
(rural: dots; urban: dashes and dots; aggregate: solid)

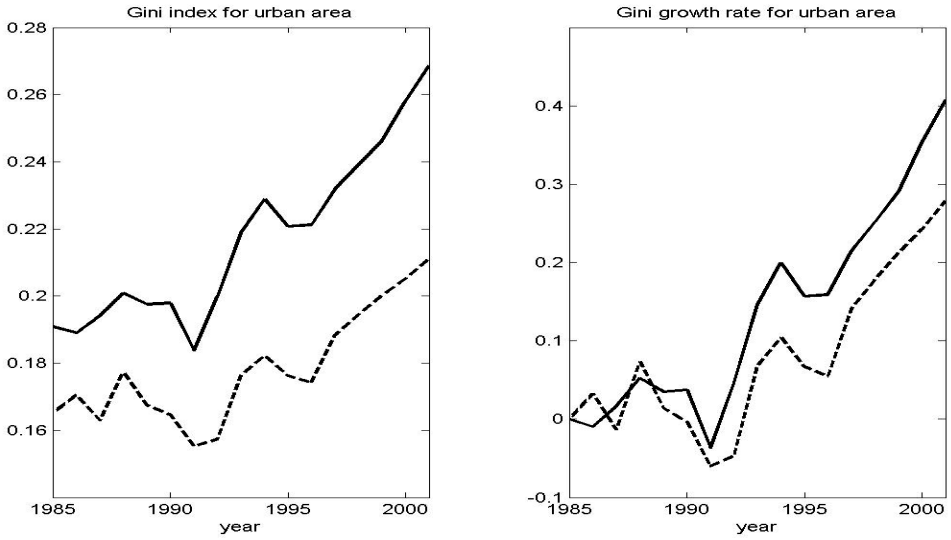

Figure 7. Gini index and growth rate for urban areas, 1985-2001
(income: solid; consumption dashes)

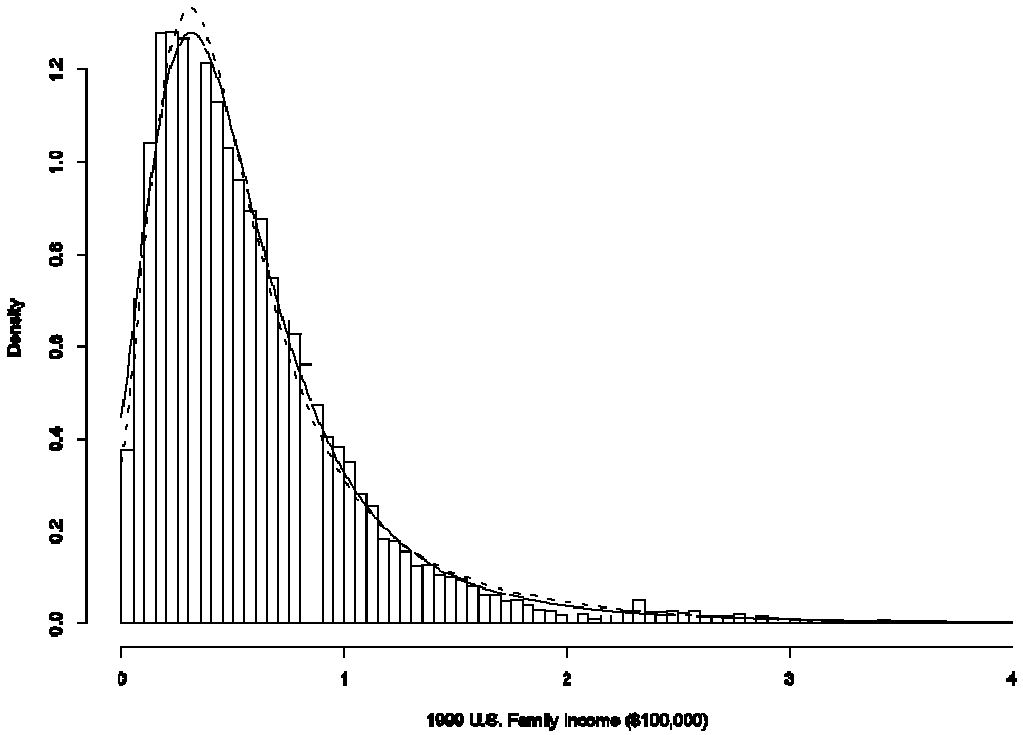

Figure A1. Estimated maxent densities (p_2: solid; p_3: dashed; p_1, which nearly perfectly coincides with p_2, is not shown)