THE WILLIAM DAVIDSON INSTITUTE
AT THE UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

The Evolution of Income Inequality in Rural China

By: Dwayne Benjamin, Loren Brandt and John Giles

The Evolution of Income Inequality in Rural China

Dwayne Benjamin
Department of Economics
University of Toronto

Loren Brandt
Department of Economics
University of Toronto

John Giles
Department of Economics
Michigan State University

This draft: December 23, 2003[*]

The Evolution of Income Inequality in China

Dwayne Benjamin
Department of Economics
University of Toronto

Loren Brandt
Department of Economics
University of Toronto

John Giles
Department of Economics
Michigan State University

Abstract

We document the evolution of the income distribution in rural China, from 1987 through 1999, with an emphasis on investigating increases in inequality associated with transition and economic development. With a backdrop of perceived improvements in average living standards, we ask whether increases of inequality may have offset, or even threaten welfare gains associated with economic reforms. The centerpiece of the paper is an empirical analysis based on a set of household surveys conducted by the China's Research Center for Rural Economy (RCRE) in Beijing. These surveys permit us to construct a set of comparable estimates of household income and consumption from a panel of over 100 villages from nine Chinese provinces. We provide a variety of summary statistics, including Gini coefficients, as well as more nonparametric summaries of the income distribution (i.e., Lorenz curves). In addition, we decompose the sources of inequality, exploring the contributions of spatial inequality to overall inequality, and the role of non-agricultural incomes in explaining rising dispersion of incomes.

We find that the distribution of income improved by most measures during the early part of the period, as average incomes rose substantially with only a modest increase in inequality. However, the distribution has worsened significantly since 1995, with rising inequality, and falling absolute incomes, especially at the bottom end of the income distribution. We attribute most of the recent decline in welfare to collapsing agricultural incomes, probably brought about by lower farm prices. At the same time, increasing non-farm incomes have widened the gaps between those with and without access to non-agricultural opportunities. Based on explorations with different data sets, our RCRE-based results probably understate the divergence due to non-agricultural income growth and the increase in inequality over time. Our results highlight the need for further evaluation of the role of farming as a source of income in the countryside, and also underline the limitations of a land-based (and essentially grain-based) income support and redistribution mechanisms.

1.0 Introduction

China has recorded impressive growth over the past twenty years, with a commensurate increase in average living standards.[1] However, there is mounting concern that increases in inequality indicate that many are being left behind, and failing to share the fruits of development.[2] Indeed, most well-behaved social welfare functions rank unequal distributions below equal ones with the same mean income levels: an important question is whether recent increases in inequality are sufficiently high to offset general rises in average income. Is rising inequality an uncomfortable, but otherwise innocuous price to pay when the rising tide is raising all boats? Not necessarily, as there is additional concern that high levels of inequality may slow down economic transition, and hinder future growth.[3]

One striking feature of the current discussion of inequality in China is the absence of well documented facts about temporal changes in the patterns and structure of the income distribution, and inequality. To some extent, this reflects lack of access to nationally representative data, so that China has been cited as an exception to global trends in our understanding of inequality in the developing world.[4] Certainly, the level of basic knowledge about China's income distribution pales in comparison to attention paid to developed countries like the U.S., or other developing countries like Taiwan, Thailand, or India. Furthermore, China is simply bigger and more complicated (an increasing urban–rural gap, and gaping regional differences) so that summaries of inequality may be less meaningful than elsewhere. Still, simple impressions of inequality have emerged from the existing literature: First, inequality has gone up during the transition; second that this is largely driven by widening inter-provincial income differences; and third, that in rural areas, the development of non-farm opportunities has provided uneven rewards for households, and is an important underlying source of inequality.[5]

Our objective in this paper is to partially fill the gap in basic facts concerning inequality in rural China. The centerpiece of our work is the use of a nationally representative household survey that has been collected continuously from 1986 to the present, covering most of the reform period.[6] By using a common household survey across years, we are able to address a number of important methodological

---

[1]Annual series from the NSB suggest an average GDP per capita growth rate 8.2 percent from 1980 and 2000 (NSB, 2003). Rawski (2001) among others have criticized China's recent GDP statistics, but few dispute the considerable growth that China has experienced since the onset of reforms.

[2]The World Bank (2003) Beijing office notes concerns about the consequences of increasing inequality for support for continued economic reform, and the new leadership in Beijing openly voices concerns about potential adverse consequences of rising inequality (Hutzler, 2003).

[3]See Aghion, Caroli and Garcia-Penalosa (1999) for a review of the growth-inequality relationship from the perspective of new growth theory; Banerjee and Duflo (2000) for a cross-country growth regression analysis; and Dayton-Johnson and Bardhan (2002) for an analysis of the impact of inequality on common property management in village settings.

[4]See Angus Deaton (2003).

[5]See, for example, Gustafson and Li (2002), Xin Meng (2003), Morduch and Sicular (2002), Rozelle (1994).

[6]There were gaps in the panel in 1992 and 1994 when the survey was not conducted.

and measurement issues associated with estimating the extent of inequality at a point in time, and comparing inequality across time periods. By employing additional, different data sets, we are also able to explore the sensitivity of our conclusions to the use of our primary data set. We also provide simple decompositions of inequality by space (village and province) and source of income (e.g., farm and non-farm) that yield important insights about the evolution of inequality.

Ideally, we would like to attribute changes in inequality to various factors associated with economic transition (moving to a market economy) and development (e.g., growth of a non-farm sector).[7] This is difficult, however, as both processes are potentially confounded in a common trend, and almost certainly intertwined anyway. Moreover, some of the recent rise in inequality appears to be the by-product of collapsing agricultural prices, and not the consequence of a "Kuznets-like" structural process.[8] While far from perfectly integrated with world markets, crop prices in China fell by more than a third between 1996 and 2000, mirroring large drops in world prices.[9] Since many households in rural China are still dependent on crop production for a substantial portion of their incomes, they have experienced absolute, not just relative declines in their standard of living. While we cannot explain the drop in crop prices, we show that it is clear that if prices remain low, without off-setting rapid development of the more equalizing sources of non-farm income, many in the countryside will remain poor, with commensurate political and migration pressures. In fact, we may now be observing China in a state of transition, with low returns to agriculture (as currently structured) a catalyst for a secular shift away from farming. Whether the Chinese economy is flexible enough to facilitate this transformation, or public policy sufficiently nimble and focused on the human cost of this potentially massive adjustment, is an obvious question.

An outline of our paper is as follows. We begin with a review of the existing literature on rural inequality, highlighting data and conceptual issues that we focus on in our work. We then describe our main data set, based on a panel of villages surveyed by the RCRE.[10] Included in this description is a summary of how we define and measure income and consumption consistently across time periods between 1987 and 1999. We then summarize various features of the income distribution for selected years in this time span. Specifically, we show results for average incomes, and a variety of measures of income inequality. We show that while average incomes have undeniably risen, so has inequality. Most disturbing, we show that after initially rising, the absolute living standards of the poor declined considerably from 1995 to 1999, so that they approach income levels of 1987: the rising tide did not lift

---

[7]This simple (and optimistic) two-factor attribution of the sources of inequality during Chinese transition is outlined in Benjamin and Brandt (1999).

[8]For the original discussion, see Kuznets (1955).

[9]As an example, the price of rice in the United States fell from $463.97 per metric ton in 1996 to $367.36 per ton in 2000, a nominal decline of about twenty percent.

[10]RCRE refers to The Research Center for the Rural Economy at the Ministry of Agriculture, Beijing.

all boats. After describing the overall distribution, we then present a spatial decomposition, where we challenge the perception that inequality is primarily a geographic phenomenon. In fact, most inequality is local. Our final exercise breaks down total income by source, where we see that the increase in inequality is driven by the combination of falling farm incomes with rising non-farm incomes. In the final section, we offer some interpretation of these results, and outline questions for future research.

2.0   Previous studies, focusing on data issues

There is an extensive literature concerned with the evolution of inequality in rural China since the onset of the reform in the late 1970s. The primary focus has been on: i) estimating the level of inequality and its changes over time; ii) and identifying the underlying sources of the inequality and its changes. At the risk of some simplification, the general consensus is that inequality has increased significantly over time. Motivated by the contrasting economic performance of rich and poor provinces, much of this increase is attributed to spatial differences linked to regional factors, and the highly uneven rate of growth of the non-agriculture sector in the countryside.

Before reviewing existing studies and presenting our own results, it is important to discuss a variety of data and methodological issues relevant for comparisons of inequality over time.[11] Imagine that the distribution of income for the population of N individuals in rural China at a point in time can be represented by:

$$f_t(x) = f_t(x_{1t}, x_{2t}, x_{3t}, ..., x_{Nt})$$

A common summary of the inequality implicit in this distribution is the Gini coefficient, given by:

$$G(x) = G(x_{1t}, x_{2t}, x_{3t}, ..., x_{Nt})$$

One cannot simply compile Gini coefficients from different sources and years in order to establish a trend. Key issues that need to be addressed in constructing informative, comparable Gini's include: The definition and measurement of the underlying welfare indicator, x_i; the unit of observation of the data; the sampling frame; and the temporal nature of the data.

2.1 Measuring welfare

The two most commonly used measures of welfare are consumption and income, and there are a variety of issues concerning the choice between them.[12] Obviously, inequality comparisons need to be made with a common welfare indicator, and there are good reasons to want to see summaries based on

---

[11]See Deininger and Squire (1996) for an excellent overview of general issues pertaining to comparisons of inequality over time (and across surveys).
[12]See Deaton (1997) for a detailed discussion of these issues.

both. That said, there are a number of issues that affect any selected measure, especially in the context of an economy where markets are imperfectly and unevenly developed.

a.   Valuing self-supplied consumption

Although declining over time, a significant portion of rural household activity is directed towards production for the household's own consumption. Thus, a significant portion of the households' income will represent the implicit return to the production on their own behalf. For household data collected by China's National Statistical Bureau (NSB) in the 1980s, however, self-supplied consumption and the revenue from these activities was valued at prices that are well below market-clearing levels. In 1992 (and beginning with data for the calendar year 1990), the NSB began using a new set of imputed prices that were significantly higher than the old set. However, the average imputed price for households' own consumption remained well below the market price, and, in fact, was only roughly equal to the farm quota price.[13] This results in an underestimate of the household's welfare and will bias our estimate of changes over time.[14]   Chen and Ravallion (1996) and Ravallion and Chen (1999) expend considerable effort correcting for these deficiencies in their sub-samples of NSB data, and show that all credible estimates of inequality in rural China must take this pricing issue into account.

b.   Durable consumption (income)

A second issue concerns the valuation of services from household durables, including housing. This component of both consumption and income (imputed rental value) is not included in NSB measures, but is incorporated (for example) in the revisions by Khan and Riskin (1998) and Khan, Griffin and Risk (1999). This results in an upward revision of household consumption and income figures. The impact on inequality of taking housing and durables into account can be significant. For example, housing services may be relatively equally distributed at the beginning, but more disequalizing with different rates of accumulation of housing. A similar case may hold for durables, which comprise an increasing share of consumption as economies develop. Especially as durables and housing may reflect consumption decisions linked to permanent income (more than, say, food), it is important to consider how they are treated in the construction of consumption (and income if imputed rental values are included in income).[15]

c.   Adjusting for Spatial Price Differences

A third issue relates to adjusting these measures for price differences. It is customary to deflate time-series data using annual price deflators in order to convert incomes and consumption totals into constant dollars (yuan). Yet spatial differences may also be important. Prices in some parts of the country may be systematically higher or lower than those elsewhere, in which case our estimates of inequality can be biased. For example, if price levels and incomes are positively correlated so that prices are higher

---

[13]The farm quota price is the price farmers received for mandatory produce (typically, grain, cotton and oilseed sales to the state.) Up through the late 1990's these prices were below the market level.
[14]The bias largely arises from the fact that the self-supplied component is declining over time so that more of farm output is being valued at "market" prices. The ratio of the market to quota price was also not constant over time.
[15]Since the rental market for housing in rural China is very limited, we do not impute rental values of housing as income in this study.

(lower) in higher (lower) income areas, spatially-unadjusted measures will lead to an overestimate of inequality of living standards (as opposed to nominal incomes). Chen and Ravallion (1996) and Tsui (1998) both construct spatial deflators for their analysis, albeit, for only four and two provinces, respectively.

2.2 Underlying Unit of Measurement

In order to estimate inequality, measures of *individual* welfare are used. As individual consumption data are rarely available, and individual income is probably an inappropriate concept in household-based production, individual living standards (income or consumption) are calculated on the basis of some per capita estimate. Usually, these are obtained from household-level surveys that provide estimates of household income, consumption, and household size. Per capita estimates are then imputed to all members within a given household, and used to estimate overall inequality. However, in some of the analyses on rural China, the chosen per capita measures have been based on units of measurement at higher levels of aggregation than the household, e.g. the village, township, county or province. These analyses effectively impute to every individual in the village, township, or larger unit, the same per capita measure. Although useful for some purposes, this has two consequences for the study of income inequality. First, it underestimates inequality because it ignores the inequality arising from differences in welfare among households within these administrative units.[16] Second, such data form a poor basis upon which to evaluate the contribution of cross-unit inequality to the total. By construction, overall inequality calculated this way will be based on differences in mean incomes between these units, e.g. differences in mean incomes between regions or provinces. Within-unit (i.e., within village or within-province) inequality is set to zero. This can also limit the types of questions that can be posed. For example, the level of township and village enterprise (TVE) development has been identified as an important source of cross-county differences in incomes.[17] But, if cross-county differences are only a small component of overall inequality, TVEs by themselves are not overly important for explaining inequality.

2.3 Sampling Frame

In order to compare levels of inequality from a country as vast as China, it is important that the samples be representative of the same population, and ideally, all of China. In principle, the National Statistical Bureau's (NSB) annual rural household survey, which dates from the early 1980s, provides

---

[16]The same problem applies when imputing every member of a household the household's mean income (consumption). On this point, see Haddad and Kanbur, (1990). Cai, Wang and Du (2002) is an example of studies of cross regional inequality using aggregate data.

[17]As in Rozelle (1994).

such a survey.[18] Although the NSB annually reports Gini coefficients based on the full sample, in-depth analysis of NSB data has been limited to select sub-samples of provinces, and then only for a few years (Chen and Ravallion (1996); Ravallion and Chen (1999) ; Tsui (1998)). The full sample NSB household data have not yet been systematically analyzed. The nearest thing to a national analysis of changes in inequality is based on a NSB sub-sample of roughly 4000 and 8000 households from 18 provinces for 1988 and 1995, which also uses a slightly modified survey instrument (Kahn and Riskin, 1998).[19] Most studies are thus based on samples with limited geographic coverage. Not only does this potentially limit the ability to estimate the impact of geographic income differences, it reduces comparability of Gini's across studies.

2.4 Time Coverage

Obviously, a trend cannot be established with one year's observation. Given the cyclical pattern of average incomes in China over the past 20 years, it is also risky to base inference of trends on even two observations. Except for Bramall's summary of NSB aggregates, there are no studies that track inequality in China on anything approximating a continuous basis. However, even for comparisons of Gini coefficients from two years, there are a number of potential difficulties. The samples need to be drawn from the same geographic units, i.e., they have to be representative of the same populations. While less important for inequality (a within-year and scale-independent calculation), it is important that adjustments for aggregate price levels be made for any welfare comparisons across time. Results may be sensitive to the chosen price deflators. Definitions of income need to be consistent across surveys. This may not be easy to accomplish, as survey designs optimally evolve to incorporate changes in the composition of income. Nevertheless, due attention needs to be placed on constructing income in as constant a way as feasible across time periods.

2.5 Findings in the Literature

The NSB data have been the sole source of estimates of long-run trends in rural income inequality using roughly comparable household-level data. While details of construction of the Gini's is sketchy, Bramall's (2001) summary of NSB trends shows an increase in the national Gini coefficient for rural China of almost fifty percent, from 0.24 in 1980, to 0.31 in 1990, to 0.34 in 1995, and finally to 0.35 in 1999.[20] The revised data of Khan and Riskin (1998), which covers a smaller sample of households,

---

[18]Unfortunately, we do not have a great deal of information or independent verification on the selection of localities and households, and on how issues of replacement are dealt with. Bramall (2001) discusses some of the potential weaknesses in the rural sampling and design, of which we say more below.

[19]An examination of the number of households surveyed in each province suggest that they reflect the share of each province in the total rural population. The 1995 and 1988 samples are not directly comparable, however, because of differences in the counties from which the households were sampled in the two years.

[20]For example, we do not know if the estimates are for household income, or per capita household income. We also do not know how the household observations are weighted. In principle, household observations should be weighted

suggest both higher levels of overall inequality and a more rapid increase over a sub-period, with the Gini coefficient rising from 0.34 in 1988 to 0.42 in 1995.[21]

Several studies have examined the spatial aspects of China's rural inequality. A majority of these, however, use provincial-level or sub-provincial per capita figures, as opposed to individual per capita household incomes. Only 4 out of 16 recent papers summarized by Gustafsson and Li (2002), for example, used household-level data. Papers using the regionally aggregated data have been interested in looking at trends in inter-provincial inequality, or have used the provincial-level data to analyze inter-regional trends. A few studies have employed household-level data to decompose rural inequality into spatial components.[22] Benjamin, Brandt, Glewwe and Li (2002) showed (for a single point in time) that within-village inequality dominated cross-region inequality, although their study had limited geographic coverage. Gustafsson and Li (2002), using the data of Khan and Riskin (1998) for 1988 and 1995, is the most comprehensive study exploring the spatial dimension of inequality over space and time. Their analysis suggests that the contribution of spatial differences at the county or provincial level was significant, and rising between 1988 and 1995.

Finally, a number of other studies (Hare (1994); Khan and Riskin (1998); Tsui (1998)) have used household data to look at the role of the emerging non-agricultural sector in explaining inequality. Such studies emphasize the potential role played by the changing structure or composition of income in generating higher inequality. With the exception of Benjamin, Brandt, Glewwe, and Li (2002), these studies do not separate or net out the spatial dimensions of income composition, but they do confirm the significant contribution of non-agricultural income sources to income inequality. Interpretation is handicapped, however, in the way that alternative income sources are often lumped together. For example, wage income from local and non-local sources are usually aggregated, while total wage income is sometimes combined with income from family-run businesses. Insofar as these sources of income are less than perfectly correlated with each other, grouping them together may hide important aspects of emerging inequality and their links to household attributes and the external economic environment with which these households interact.

3.0 Data

---

by household size (to mimic individual level data). Finally, recall that income is not consistently measured across surveys, and the underestimate of agricultural income in early years probably leads to an overstatement of initial inequality.

[21]See previous footnote, however.

[22]Benjamin and Brandt (1997) show that within-village inequality was also the dominant source of overall inequality in 1930's China, though their sample came from only two provinces.

The data used for our analyses are the product of annual household surveys conducted by the Survey Department of the Research Center on the Rural Economy (RCRE) in Beijing. Household-level surveys from over 100 villages in 9 provinces (Anhui, Gansu, Guangdong, Henan, Hunan, Jiangsu, Jilin, Shanxi and Sichuan) are matched with corresponding village-level data.[23] In each province, counties in the upper, middle and lower income terciles were selected, from which a representative village was then chosen. Subject to the limits of this stratification, the RCRE sample should reasonably capture both inter- and intra-provincial income variation. Depending on village size, between 40 and 120 households were randomly surveyed in each village. The survey spans the period 1986 to 1999, and includes between 7,000 and 8,000 households per year.

The RCRE originally intended a longitudinal survey, following the same households over time. While there is a household-level panel dimension to our sample, we observe considerable attrition of households over the 1986-1999 period, especially after years when there was no survey. The RCRE was unable to conduct the survey in 1992 and 1994 because of funding difficulties. Households lost through attrition were replaced (at least in principle) on the basis of random sampling.

The survey collected detailed household-level information on incomes and expenditures, education, labor supply, asset ownership, land holdings, savings, formal and informal access to credit, and remittances.[24] The large number of households surveyed from each village and the lengthy span of the survey enables us to track the evolution of consumption, incomes and inequality during a time of changing market access and development in rural China. Of particular importance for our purposes, we are able to track a panel of villages, even where there has been household attrition. This will allow us to maintain geographic comparability over the complete time period.[25]

A variety of definitions are worth clarifying, and further details related to attrition issues are provided in the Data Appendix. First, household membership is defined on the basis of residency and registration.[26] Second, income is calculated as the sum of net income (gross revenue less current expenditures) from agriculture, farming sidelines (e.g. animal husbandry and livestock), family-run businesses, plus wage income, and transfers. We calculate the value of farm output that is not sold, and

---

[23]The complete RCRE survey covers over 22,000 households in 300 villages in 31 provinces and administrative regions. By agreement, we have obtained access to data from 9 provinces.

[24]One shortcoming of the survey is the lack of individual level information. However, we know the number of dependents and individuals working, as well as the sex composition of household members.

[25]It should be noted that we also witness some attrition of villages. For the most part, this attrition was driven by disagreements with village leaders. When a village is lost from the sample, it is replaced with a "comparable" village chosen from the same county.

[26]It includes individuals in the household with rural registration (*hukou*) plus a small number of individuals with non-rural registration, but who live in the village full-time. This definition of household membership differs slightly from that of the World Bank's Living Standard Measurement Survey, which for other than the household head, bases membership on the actual number of months of residency in the house.

thus largely consumed (or stored) by the household itself, at market prices.[27] Household income is also gross of taxes and fees. Third, our measure of consumption includes non-durable goods expenditure, plus an imputed flow of services from household durable goods and housing.[28]

We deflate all income and expenditure data into 1986 prices using the NSB rural consumer price index for each province. For some key results we explore the sensitivity to geographic differences in price levels. In those cases we spatially deflate using a cross-province CPI deflator constructed by Brandt and Holz (2003), based on expenditure weights from the NSB rural household survey. The spatial CPI adjusts for systematic differences in price levels across provinces (at a point in time), because price levels and incomes are positively correlated, possibly exaggerating differences in living standards across regions. Finally, RCRE's sampling is not proportional to provincial population. For example, the number of households surveyed in Sichuan is nearly the same as that surveyed in Gansu, despite the fact that Sichuan has a rural population that is nearly 5-times larger. Therefore, we use provincial rural population (by year) as weights in our calculations.

In order to establish the robustness of our conclusions to various permutations of sample selection, we carried out our analysis on three different data sets. The first, or "full", sample includes every household (panel, attritted, and replacement) in each survey year. The second accounts for the fact that inequality measures may be sensitive to outliers (at both ends of the distribution), and this "trimmed-full" sample drops extreme outliers among households.[29] The third sample is a "balanced-panel", comprised of those households for which we have data for every year of the survey. As it turns out, our results are consistent across all three data samples. To minimize tables, we restrict our reported results to those from the "trimmed-full" data set.

4.0 Results

4.1 Income Distribution over Time

We begin by summarizing the evolution of average welfare levels over the span of our sample. Results are reported in Table 1. To keep the tables manageable, we report results for four evenly spaced

---

[27]RCRE's surveys follow NSB conventions and value non-marketed grain at quota prices. We follow an approach used in Chen and Ravallion (1996) when we recalculate the value of non-marketed grain at marketed prices.

[28]In order to convert the stock of durables into a flow of consumption services, we assume that current and past investments in housing are "consumed" over a twenty-year period, and that investments in durable goods are consumed over a period of 7 years. We also annually "inflate" the value of the stock of durables to reflect the increase in durable goods' prices over the period.

[29]The lowest income households actually have negative incomes. These are typically households that have high gross incomes, but also high business-related expenses. The problem of measurement error for these households is especially severe. We discuss our procedure for eliminating outliers in Appendix Section 7.3. In each year, fewer than one-tenth of one percent of households were dropped from the panel.

years – 1987, 1991, 1995, and 1999 – which fully reflect the patterns in the more complete sample.[30] Mean household per capita incomes were 578 (RMB) in 1987, the beginning of our selected years. Average incomes dropped slightly through to 1991, sharply rising to 772 in 1995. The average annual growth rate over the 1991-95 period was an impressive 5.3 percent, reinforcing the optimism of economic reforms after the 1989-1991 retrenchment. However, this growth was short-lived, and by 1999 average per capita incomes had actually fallen to 714. Setting aside the cyclical variation in growth, the average rate of growth from the beginning to the end of the sample was 1.8 percent. An important question is whether the decline in average incomes was disproportionately borne by the poor, as this would certainly worsen the distributional consequences of rising inequality.

Our results for consumption closely mirror those for income, both in terms of the cyclical patterns, and the implied growth rate over the entire period. Given that the only overlap in the two series is home-produced consumption, it is reassuring that these two otherwise independent measures of welfare should track each other so closely (though this may be less assuring for those who believe that consumption should be much smoother than income). In levels, consumption is approximately three-quarters of income. Some of this gap probably reflects measurement error, but it is also due to genuinely high rates of savings, and the fact that incomes are measured before deduction of taxes and other fees.

In Table 1 we also show results for spatially deflated mean income and consumption. As we can see, spatial deflation makes no difference for this exercise. This is not surprising, as the aggregate numbers do not provide much scope for differential provincial price levels to affect the evolution of the average. The deflator is more likely to matter when we compare incomes across regions, as we will see when we examine inequality.

How do the RCRE numbers compare to other data sources from China? The only other data set that spans this period is the nationally representative – and publicly unavailable – NSB rural household survey. In Figure 1 we plot average incomes for each year of the RCRE, alongside the corresponding NSB rural averages as reported in yearbooks. We do not make this comparison assuming that the NSB data are the gold standard, as the NSB surveys were not conducted with the same degree of comparability as the RCRE surveys. For example, as discussed in Section 2.1, the early NSB incomes were calculated by valuing household consumption of own-produced farm output at quota prices instead of market prices. This serves to generate lower incomes for those years when the market price significantly exceeded the quota price. That said, the NSB data are the most common, nationally representative data that can otherwise be used for this type of analysis. Nevertheless, the basic patterns in the RCRE and NSB data are

[30]Results for Tables 1 and 2 (our main results) for the complete sample of annual observations, and for the different samples (panel and non-panel) are available from the authors by request.
[32] See Deininger and Squire (1996) for a discussion of the evolution of inequality over a broad range of countries.

quite similar, with the notable difference that the RCRE starting incomes are much higher than those of the NSB. Some of this difference is due to the home-production valuation problem. The higher relative incomes of the RCRE erode by 1999, where the mean incomes in the two surveys actually converge. Thus, the RCRE data shows a flatter time series, and a correspondingly lower growth rate. Without a detailed forensic comparison of the two data sets, we can only conjecture on the reasons why the two series have the different starting points, and such a comparison is not possible given the inaccessibility of the NSB household-level data. One possible source of differences is that the RCRE provinces are not nationally representative. However, when we include only the RCRE-subset of provinces in the NSB data (which come at the provincial level), we see that there is almost perfect tracking of the RCRE-subset of provinces with the national total. Despite the differences in the magnitude of the trend between 1987 and 1999, both series agree that growth was fastest in the 1991-1996 period, with a significant attenuation (or decline) from 1997 to 1999.

Table 2 provides measures of income and consumption inequality emphasizing a variety of distributional characteristics. We begin with the Gini coefficient for income, arrayed in the first row. The Gini increased from 0.32 in 1987 to 0.37 in 1999, an increase of 0.05 or 16 percent. Is this big? There are few benchmarks of comparison, though it is worth noting that inequality measures evolve slowly over time, and a 16 percent increase is large.[32] Of particular note, almost all of this increase was over the short period between 1995 and 1999. Combined with the decline in average incomes, it should come as no surprise that concerns over inequality have intensified over the 1990's. In the second panel we show the results for consumption, where the Gini rises from 0.25 to 0.31 over the complete sample. As is usually the case, the Gini for consumption is lower than income, but the trend and over all time-pattern is basically the same. If anything, the increase in consumption inequality was slightly larger in percentage terms. Measured either way, it appears that inequality went up over this period, especially since 1995.

In Figure 2 we place the RCRE results beside those from the NSB, where the NSB Gini's are drawn from Bramall (2001). The NSB numbers show inequality rising from 0.24 in 1980, to around 0.3 in 1990, slightly lower than in the RCRE sample, and finishing off at 0.35 in 2000. As with the RCRE data, the NSB data thus show the Gini rising by about 0.05 points over the 1990's. While there are slight differences in the magnitudes of the Gini's in particular years -- which can potentially be attributed to differences in sampling, sample sizes, stratification, or other differences in the survey instrument – both the trends and magnitude of the increase in inequality are similar in the two surveys. Thus, while the NSB and RCRE data differ in detail, they paint a similar overall picture for the evolution of rural household welfare over the 1990's.

Returning to Table 2, in the second row we apply the spatial price deflator to household income. As expected, the magnitude of the Gini drops. Higher income areas appear less well off once due account

is taken of the higher prices faced by consumers in these provinces. The magnitude of the drop is 0.02 to 0.03, and spatial deflation does not materially affect our conclusions about the overall trend in inequality: In other words, we cannot deflate away the increase in inequality and attribute it to widening gaps in regional nominal (versus real) incomes.

While the Gini shows an increase in inequality, is there any sensitivity of our conclusions to our particular choice of inequality measure? The next four rows of the table present inequality measures for the variance of log income and the Atkinson index evaluated at three different inequality aversion parameters (ranging from 2.0 to 0.5 in decreasing magnitude of inequality aversion). While these measures can be used for comparison to other studies, the main purpose they serve for us is to confirm that inequality went up, especially between 1995 and 1999. Most worrying, the Atkinson index with high inequality aversion increases from 0.33 to 1.00 over this four year period! This suggests that the bottom part of the income distribution did especially badly.

Of course, direct comparison of the Lorenz curves yields a non-parametric comparison of potential changes in inequality, possibly avoiding the need to choose any of the inequality measures in Table 2. We show the Lorenz curves for 1987, 1995, and 1999 in Figure 3. Panel A shows the curves for the entire sample, while panel B provides a close-up view of the bottom part of the distribution (from Panel A). Most striking is that the 1999 Lorenz curve lies (well) outside those for 1995 and 1987, suggesting that for any inequality measure we choose, 1999 will look worse than those years. Unless there was a corresponding increase in average welfare in 1999 (which there wasn't), no conventional social welfare function would rank 1999 above 1995 or 1987. In terms of inequality, 1987 and 1995 are almost indistinguishable. Unambiguous comparisons are not possible, as these Lorenz curves cross, with a slight advantage (in terms of inequality) at lower incomes for households in 1987.

A simple way to incorporate average incomes into the Lorenz curve is through the Generalized Lorenz Curves shown in Figure 4. Basically, the Lorenz curves from Figure 2 are multiplied by average annual per capita income. A Lorenz curve illustrates the share of the pie going to lower income percentiles, while the generalized Lorenz curve adjusts for the size of the pie. In Panel A we see that the curve for 1995 lies strictly above the other years, indicating that the distribution from 1995 Lorenz dominates those from 1987 and 1999. At the top end of the distribution, the 1999 distribution is second best, as richer households are better off than those from 1987. What is most striking, however, is that for the bottom quarter of households, the 1999 curve is barely indistinguishable from 1987. The Generalized Lorenz Curves cross around 0.18, with 1999 below 1987 for the bottom 15 percent of individuals. The absolute size of the pie going to the bottom 10 percent of the population (for example) is smaller in 1999 than 1987. This is consistent with the poorest households being no better off after 12 years of economic growth! If true, it is no wonder that there is concern about inequality: the rising tide is not lifting up all

boats. Even worse, the bottom part of the distribution has lost this significant ground in four short years. Perhaps this stark result is sensitive to the particular price deflators we use (the conventional Provincial CPI's from the yearbooks). However, we doubt that the deflator actually makes much of a difference in this case, except to the extent that cheaper food prices may not be fully reflected. That said, even if the deflator is imperfect, it is troubling that whether households are better off in 1999 than 1987 should come down to details of CPI measurement.[33]

In the next two rows of Table 2 we switch from overall measures of inequality to considerations of relative inequality between rich and poor, and poverty. First we present the proportion of households with incomes below half the contemporaneous mean.[34] Essentially, this amounts to setting a "relative poverty line" at fifty percent of the mean income of that year: this poverty line moves with average incomes and while it does not tell us much about absolute poverty, it provides another useful way of characterizing the distribution of income. The proportion of the sample below half contemporary mean income increases from 16 percent in 1987 to 22 percent by 1999, again with most of the jump occurring after 1995. These results reinforce our conclusion that the increase in inequality reflects a worsening of the relative position of low income households: using this relative measure, the rich are getting richer and the poor are getting relatively poorer. A stronger point about absolute changes in poverty can be made when we keep the "poverty line" constant at half 1987 mean levels (in real terms). We can thus track progress on the elimination of poverty with a constant benchmark. Here we see a worsening of the poverty rate between 1987 and 1991, which comes across as the worst year for the poor. The best year was 1995, with half as many people "poor" as in 1987. By 1999, however, the poverty rate (so-measured) has returned to essentially the same level as 1987 (and doubling in four years from 1995). Despite an increase of average incomes by 25 percent between 1987 and 1999, the fraction of people below the 1987 "poverty line" has barely budged. No wonder perceptions of rising inequality are a matter of great concern.

As a final exercise in the poverty dimension, we plot the cumulative distribution functions (CDF) for 1987, 1995, and 1999. This exercise permits using any common poverty line applied to the three years to compare poverty rates. For example, if we chose the 1987 half-contemporaneous mean of 290 RMB per capita as our poverty line, we can recover the poverty rates from Table 2: approximately 6 percent of people have incomes below 290 in 1995, compared to 14 percent in 1999 and 16 percent in 1987. Because the 1995 CDF lies everywhere below those for 1987 and 1999, the distribution from 1995 is stochastically

---

[33]The issue is whether cheaper food prices are fully reflected in the rural CPI. Since rural CPI weights are based on *purchased* food, to the extent that consumers have cheaper home-produced food, their increased purchasing power may not be fully captured by the CPI.

[34]See Jenkins (2000), for an example of using the "half-contemporaneous mean" in summarizing the evolution of the distribution of income in the UK.

dominant. The CDF's for 1987 and 1999 cross at about 200 yuan, so that for poverty lines at or below 200, the fraction of poor people is higher in 1999 (than 1987), but for higher poverty lines, the 1999 fraction of poor is lower. The CDF's can be used to highlight the fact that the bottom poorest 20 percent are only marginally better off in 1999 than they were in 1987, while the distribution of income is unambiguously favorable for higher income households in 1999 (compared to 1987), at least for those with per capita incomes above 1000 (the top 15 percent).

The final row of the top panel of Table 2 reports the ratio of the 90th to the 10th percentiles of the income distribution, providing another angle on the widening disparities of income. In 1987, the ratio of rich (90th percentile) to poor (10th percentile) per capita incomes was 4.06. This ratio was approximately constant through to 1995, but increased significantly by 1999, where "rich" individuals were more than five times richer than the "poor." As we saw in the previous discussion, this was driven more by a collapse of incomes at the bottom than an explosion of incomes at the top.

Finally, the bottom panel of Table 2 shows the matching results for consumption. As with the Gini's and mean incomes, the picture based on household per capita consumption is essentially the same as the one based on income.

Taken on their own, these results from the RCRE surveys are alarming in terms of the evolution of the income distribution in recent years. But in the broader context of world income inequality, a Gini rising from 0.32 to 0.37 over such a dynamic period may not seem so dramatic, at least in proportion to the concern expressed. Can the 0.37 Gini for 1999 be taken as a reliable upper bound for inequality in rural China? As we already saw in Figure 2, the RCRE results line up with the NSB. However, as we now show in Table 3, this conclusion is premature, as there are other data sets that show higher levels of inequality, and simultaneously point to potential weaknesses in both the NSB and RCRE survey designs.

For select years, we compare mean incomes, the composition of income, and inequality using the RCRE and two other household surveys. In Table 3 we report a comparison of tabulations of data from the fourth wave of the China Health and Nutrition Survey (CHNS) that covers the calendar year 1997; the RCRE survey for 1997; similar RCRE data for 1999; and a collaborative household-level survey carried out in 2000 covering 1200 households in 6 provinces (the CCAP 2000 survey).[35] We break income down into that from farming, farm sidelines (forestry, livestock, and fisheries), wage income, family-run businesses, and a residual category, "other", which is largely comprised of private and public transfers. We report total per capita income, per capita income by source, and the percentage of households in each of the surveys that report non-zero income from each source. For the CCAP 2000 survey we report two

---

[35]The 2000 survey was a collaborative effort involving Bai Nansheng (formerly of the RCRE), Loren Brandt, Scott Rozelle (UC-Davis), and Zhang Linxiu (Chinese Center for Agricultural Policy, or CCAP). See the Data Appendix for further details.

tabulations: (1) based on the full sample; and (2) based on the full sample, but dropping the top one percent of households (in terms of per capita income). In the calculation of these summary measures, we effectively weight observations in each province by that province's share in the total rural population of the sampled provinces. However, there is only limited overlap across the surveys in terms of the provinces sampled. As we saw in Figure 1, the RCRE provinces are not peculiar compared to national averages, but there will still be limits to comparability of the other surveys.

First note the comparison of the RCRE with the CHNS for 1997. The structure of income (mean incomes by source) is similar in the two surveys, with slightly higher (by less than 10 percent) income in the CHNS. To some extent, this reflects a slightly higher fraction of suburban households in the CHNS rural sample (notice the slightly smaller proportion of farmers in the CHNS). Most notable, however, is the higher level of inequality reflected in the CHNS, with a Gini of 0.43.

The most striking gap is between the 1999 RCRE and the CCAP for 2000, with average incomes higher by 22% in the CCAP survey. Income growth was flat (or negative) between 1999 and 2000, so the difference cannot be attributed to economic growth. Moreover, the rural CPI was falling over this period, and so the differences in real income are even slightly more pronounced than the nominal figures that we report. Much higher reported income from family run businesses in the CCAP 2000 data appears to be the source of most of the difference. Mean per capita income from family businesses was 796 in the 2000 survey, but only 494 for the 1999 RCRE. This difference represents 62.5 percent of the gap in mean incomes between the two surveys.

A comparison of inequality measures based on these surveys reveals an even more substantial difference in the two surveys, with the CCAP survey suggesting a Gini of 0.50, which is much higher than the RCRE (or other surveys). But a comparison with a slightly "trimmed" version of the CCAP 2000 survey identifies the likely source of the problem. Official surveys often exclude the richest households. By design, they are also less successful in accurately estimating household incomes from family-run businesses. For example, both the RCRE and CHNS surveys simply ask respondents for total revenue and expenditures from family businesses, instead of a detailed enumeration of costs and revenue, with some attempt to distinguish between fixed and variable inputs. In general, this is the most difficult source of income to enumerate accurately, and thus often the noisiest. It is also believed to drive rural inequality, so under-estimation of this income source may significantly lead to understatement of the level of inequality. Considerable effort was placed in the CCAP 2000 survey on minimizing problems stemming from inaccurate reporting of non-farm business income, both in terms of random sampling of households within villages, and a more careful enumeration of the balance sheets of family businesses.

The most striking comparison is made between the trimmed and untrimmed versions of the CCAP sample. Dropping the top one percent of households drops mean household per capita income

from 2667 to 2370, or eleven percent, and bringing it closer in line with the RCRE and CHNS estimates. Incomes by source also line up very well between the RCRE and CCAP trimmed samples. Almost all the drop in mean incomes, and resulting improvement in correspondence of the surveys, comes from the decline in average incomes from family-run businesses. The Gini coefficient also falls considerably, from 0.50 to 0.44, more in line with the inequality reflected in the CHNS. If we drop the top three percent, the Gini would fall slightly more, to 0.42.

This exercise shows several important points. First, the level of inequality (Gini) is highly sensitive to the top tail of the income distribution. To the extent that the RCRE (or NSB) surveys miss the very richest households (possibly because they are not "representatitive"), overall inequality will be understated. Second, poor measurement of family-run business income alone can lead to a significant misrepresentation of the level of inequality. The CHNS and CCAP surveys thus suggest that NSB and RCRE-based estimates of the level of inequality are too low (possibly by as much as 0.10 Gini points). To the extent that family-run businesses have been increasing in importance over the reform period, the RCRE and NSB likely understate the upward trend in inequality. The results in Table 2 are probably a lower bound of the extent to which inequality has risen. Combined with the more reliable (and consistent) results from the RCRE for the bottom part of the income distribution, we conclude that that the distribution of welfare has significantly eroded, especially over the last half of the 1990's.

4.2 Decompositions By Geography

The role of widening regional income differences and their contribution to increasing inequality is a common theme in the literature on inequality in China.[36] Rising disparities between localities, especially provinces (inland versus coastal, for example), are often seen as being the most important source of the rising income differences, as some provinces are better situated to take advantage of market liberalization and new off-farm opportunities. Our sample, which includes the rapidly growing coastal provinces of Guangdong and Jiangsu, but also the interior provinces of Sichuan and Gansu, seems reasonably well suited to look for these trends.

There are a number of approaches one can take in decomposing inequality along dimensions, like region. Unfortunately, the Gini coefficient is not readily (or neatly) decomposed. Gustaffson and Li report spatial decompositions for the decomposable Mean Log Difference and Theil inequality indices. We adopt a simpler strategy, decomposing the variance of log income inequality index. This entails estimating the following regression:

---

[36]Kanbur and Zhang provide an excellent overview of the literature on regional inequality, highlighting inland versus coastal, and urban versus rural dimensions. Gustaffson and Li (2002) provide an especially helpful and detailed summary of the literature on spatial inequality. See in particular their "Table 1" summary of the literature, including a catalogue of the data sources used by various authors.

$$\ln y_i = D_{Li}'\gamma + u_i$$

where D_L is a vector of dummy variables indicating the location of individual i. The R-squared from this regression indicates the proportion of the variation (or variance) of $\ln y_i$ that is explained by the location dummies, and this is our estimate of the fraction of inequality explained by location. The remainder is the (within-location) residual variance of log income, and a measure of the degree to which household income cannot be explained by the average income of its neighbors.

Table 4 reports the results of this exercise, conducted along several dimensions. The top half looks at income inequality, while the bottom panel reports the results for consumption. We define location at three levels of aggregation. The building block is the village, and is the finest level of aggregation. We then move up to the provincial level, with the nine RCRE provinces, and ultimately region: West (Gansu, Shanxi, and Sichuan), Central (Anhui, Henan, and Hunan), and East (Jilin, Jiangsu, and Guangdong). Our final cut of the results distinguishes between spatially deflated and undeflated household incomes, which assesses the extent to which regional income differences are attenuated once we take account of spatial differences in the cost of living: a dollar of income may not buy the same amount of welfare in each province.

In the first row of Table 4, we see the proportion of inequality explained by region declined from 0.19 in 1987 to 0.15 in 1995, and to 0.12 in 1999. While there are limits to comparability, we discuss our results in light of Gustaffson and Li's. Their results for the proportion of Mean Log Difference (MLD) (which is similar to the variance of logs) explained by region are 0.12 for 1987, and 0.27 in 1995. Our results thus differ in both level and trend. Some of this difference may be driven by the different decomposition procedures, as well as differences in measurement of income. While the provinces are not the same in the two studies, the regions are similarly defined. Another limitation of comparison is that the smallest unit of analysis for Gustafsson and Li is the county, not the village. Furthermore, while we essentially have a panel of villages in the RCRE (87 out of approximately 110 villages are the same across time periods, and the base counties are the same), they do not have a panel of counties. Differences in counties across time periods may add another dimension of non-comparability across time for studies based on Kahn and Riskin's sub-sample of NSB data (as with Gustafsson and Li).

When we turn to province-level results, the difference in patterns between studies is also apparent, though less pronounced. Here, we find the proportion of income inequality explained by province declined from 0.24 to 0.18 between 1987 and 1995, and further to 0.15 by 1999. By contrast, Gustafsson and Li start out with a similar proportion explained by province in 1987 (0.24), but their proportion rises to 0.32 by 1995.

In Figure 6, we explore possible differences between the RCRE and NSB household surveys (the basis of the samples used by Gustafsson and Li). Our objective is to compare the amount of province-level inequality present in the two data sets. For the RCRE, we calculate overall Gini-coefficients assuming everyone in a province earns the same income (provincial mean income). The national Gini is constructed on the basis of these inter-provincial income differences (weighted by rural population). For the NSB, we use provincial mean incomes from Statistical Yearbooks of China to calculate similar national Gini coefficients, weighting mean incomes by provincial rural population, and effectively attributing everyone in a province the same income. This procedure sets within-province inequality to zero, and calculates the implied Gini arising from differences in provincial mean incomes alone. The basic pattern in Figure 6 is similar to the one we saw in Figure 2: the level of inequality is almost the same in the two surveys, but the slope is slightly steeper in the NSB. The NSB shows an especially steep increase in spatial inequality between 1988 and 1995 (Gustafsson and Li's sample years), compared to the RCRE. Both series show a significant flattening of this trend since 1995, with the NSB showing actual declines in spatial inequality from the mid-1990's to 2000. In Figure 6, we also note that the RCRE subset of provinces had a slightly lower level of spatial inequality than the full national sample, but not enough to render the RCRE provinces unrepresentative. In summary, while there is disagreement about the initial level of spatial inequality (as we saw in Figure 2), the two data sources essentially agree on the magnitude of spatial inequality, and especially the relative decline since the mid 1990's.

Returning to the RCRE data source, our results suggest that the role of provincial income differences has declined over time. Another way to illustrate this point is to look at the behavior of provincial mean incomes and within-province inequality. In Figure 7 we plot average provincial growth rates against the initial (1987) level of income. There is no obvious pattern here, and results are sensitive to a single observation, notably Guangdong. Excluding Guangdong, this figure suggests some degree of income convergence, with poorer provinces growing more rapidly over the 1987-1999 period. This is consistent with a narrowing of inter-provincial inequality. But of course, we cannot just exclude Guangdong in painting the complete picture. Nonetheless, Figure 7 provides no evidence that provincial income levels were diverging.

In Figure 8 we plot within-province Gini's from 1999 against the provincial Gini for 1987. We also show a 45 degree line in order to benchmark the inequality levels in the two years. Here we see that inequality rose in all provinces except Gansu, and in Jilin rose to over 0.40. Again, this is consistent with the decompositions which show that within-province inequality became more important between 1987 and 1999. Furthermore, excluding Jilin, we see that the plot suggests provinces with lower inequality in 1987 had higher increases of inequality to 1999, implying convergence of Gini's across provinces.

Next, in Table 4, we evaluate the impact of spatial deflation on the decompositions. Accounting for inter-provincial price differences cuts the share explained by province or region in half. This suggests that the inter-provincial income gaps overstate the differences in the standards of living across provinces, however, the bias is declining over time. (For example, the difference in the contribution of province between the spatially deflated and non-deflated series is much larger in 1987 than in 1999). Spatial deflation has a much smaller effect on the role of village in the decompositions, but this is to be expected since the spatial deflator uses provincial level prices.

Finally, in the third row of Table 4 we show the fraction of inequality explained by village. Here, we see the proportion fall from approximately 0.50 in 1987 to 0.40 in 1999: Most of the inequality in our sample occurs within, as opposed to across villages.[37] An obvious question is whether 0.50 represents a half-full or empty glass, in terms of the role played by geography. Nothing in our conclusions diminishes the fact that location is an important (perhaps the single most important) determinant of household income. Furthermore, even a diminution of the role played by village is consistent with persistence of low incomes within and across villages, such as one would expect with geographic poverty traps as identified by Jalan and Ravallion (2002). Instead, we view our results as pointing to the significant role played by within-village differences in incomes as a contributing factor in overall inequality, and correspondingly draw attention to those factors that generate inequality within villages.

In Figures 9, 10, and 11 we explore the evolution of village-level inequality. First, in Figure 9 we show that there is evidence of convergence of income levels across villages (poorer villages tended to grow more rapidly between 1987 and 1999). All else equal, this convergence reduces the role of village in explaining inequality (as we saw in Table 4). Figure 10 shows histograms for the village-level Gini's, clearly showing the shift upwards of within-village inequality. This shift is also readily apparent in Figure 11, where we see that a majority of villages experienced increases of their income Ginis, with a considerable fraction experiencing increases over 0.10, though village sample sizes are small enough to warrant a caution on placing too much stock on a single Gini. As with the provincial-level inequality measures, it appears that there is convergence of inequality levels, whereby low-inequality villages experienced greater increases in inequality.

The broad conclusion from the spatial income-inequality decompositions is that (1) no more than half of total inequality is driven by income differences between villages, and (2) the role of geography seems to be declining, especially in recent years. In terms of understanding the sources of inequality, this should serve to turn more attention towards local institutions, and local variation in the distribution of

---

[37]Gustaffson and Li do not report results for village, as their finest unit of location is the county. With this caveat, they find that 40 percent of inequality was across counties in 1988, rising to 50 percent by 1995. While trends are opposite ours, the basic magnitude is similar. Also note that the NSB data show a decline in inter-provincial inequality from 1995 to 1999 (Figure 6), so more recent numbers from the NSB might line up more closely.

endowments (like skills, education, and land). If we use consumption instead of income, the bottom half of Table 4 shows the same basic story, though the magnitude of spatial inequality is slightly higher.

4.3 Decompositions by Source

Why has inequality gone up within villages? Answering this question requires an understanding of the evolution of institutions across villages that map household endowments into family income, and is a significant research enterprise in itself. Our more limited objective here is to sketch some of the correlates of within-village inequality, particularly those related to the composition of household income. Previous studies have emphasized the role of non-farm income in contributing to rising inequality.[38] We can use the RCRE data to confirm the role of non-farm income, to evaluate finer details of the composition of sub-components of non-farm income, and most importantly, to gauge trends in the role of income composition leading to increases in overall inequality.

The key tools in our analysis are descriptive statistics of the structure of income, and Shorrocks (1982, 1983) decompositions. The Shorrocks decomposition is designed to answer the question, what proportion of total inequality can be attributed to inequality of income source k? It is a purely descriptive tool, and there are clear limits to the extent that one can attribute a causal interpretation to the coefficients. For example, it is difficult to use the results to simulate the impact of an increase in the inequality of a particular income source on total inequality, without further specifying the nature of the increase in inequality of that income source. However, as we shall see, even within the limits of interpretation the decompositions are illuminating.

As an outline of the procedure, consider a decomposition of the mean of household income, based on household i's income, y_i:

$$y_i = \sum_{k=1}^{K} y_{ik}$$

which is the sum of K sub-components of income y_{ik}. Clearly, mean household income can be written:

$$\overline{Y} = \overline{Y}_1 + \overline{Y}_2 + ... + \overline{Y}_k$$

A one-percent increase in mean income from source k will lead to a W_k proportionate increase in $\overline{Y}$, where W_k is the share of income from source k. Decomposition of the sources of mean income is thus

---

[38]See Benjamin, Brandt, Glewwe and Li (2002) for a survey of these studies, e.g., Hare (1994), Kahn and Riskin (1998), Tsui (1998), and Morduch and Sicular (2002).

straightforward, and decomposition of inequality is designed analogously. We wish to estimate S_k, the proportion of inequality attributable to the inequality of income source k:

$$I(Y) = \sum_{k=1}^{K} S_k I(Y_k)$$

where $I(Y)$ is the index of inequality for total income Y, and $I(Y_k)$ is the index of inequality for income source k. Shorrocks showed that whatever the index (e.g., the Gini or some other measure of inequality), S_k is estimated by:

$$\hat{S}_k = \frac{\text{cov}(y_{ik}, y_i)}{\text{var}(y_i)}$$

So, S_k captures the degree to which income source k is correlated with total income. In this sense, it measures the degree to which particular income sources are earned by the rich or poor. If an income source is earned primarily by the rich, then the decomposition will attribute a larger share of total income inequality to inequality of income earned from that source. How can we interpret these S_k? One benchmark is zero: if an income source is negatively correlated with total income, then it is earned disproportionately by the poor, and no inequality (indeed a negative share) of total income is correlated with that income source. Presumably, marginal increases of inequality of that source of income (maintaining the same correlation with total income) would further reduce overall inequality. Very few sources of income will have negative S_k. Another helpful benchmark is the mean share of income from that source, or W_k. If $S_k > W_k$, then inequality of income source k contributes more to inequality than it does to mean income, $\overline{Y}$, which we denote as a disproportionate effect on inequality. In other words, if income from family businesses comprise 10 percent of average income, but 20 percent of inequality, we will conclude that family business income has a disproportionate effect on inequality.

As a matter of computation, $\hat{S}_k$ can be estimated by the following regression:

$$y_{ik} = \beta_{0k} + \beta_{1k} y_i + u_{ik}$$

as $\beta_{1k} = \hat{S}_k$. This regression presentation also aids in interpretation: all we are estimating is the correlation of a particular source of income y_{ik} with total income, y_i. Once we broaden our objective to the estimation of this correlation, we can also recognize the possible impact of measurement error: overestimates of income from a particular source will lead to an overstatement of the correlation with total income, and β_{1k} will be overstated. This overstatement for income source k will spill over to the

other β_{1k}, leading to an underestimate of their contribution. One simple way to address this possibility is to estimate the regression by two-stage least squares, using another indicator for total income as an instrument for y_i. An obvious candidate is total household consumption, which should not suffer from the same type of measurement error as the y_{ik} (though it certainly suffers from its own measurement error). Aside from the instrumental variables interpretation, this procedure can be viewed as exploring the sensitivity of our conclusions concerning the correlation of income from a particular source with whether a household is rich or poor, to alternative definitions of rich and poor, based on income or consumption. As a final refinement on the Shorrocks procedure, we compare decompositions with and without village dummies. Inclusion of the village dummies allows us to decompose within-village inequality, and to net out the possible effect of variation of income sources across villages, and this link to cross-village inequality.

We begin with a description of mean incomes by source, reported in Table 5. In 1987, agricultural income (crop income) comprised 40 percent of total income. The largest sub-component was grain income, at 30 percent of total household income. Adding income from agricultural sidelines (fish, forestry, and especially livestock) raises the broadly defined share of agriculture to 53 percent, over half of family income. Family business, mostly in commerce and services, comprised 16 percent of income, while wage income was the second largest overall component, at 25 percent. Most wage income was earned locally, within the village.

This structure of income changed dramatically by 1999. Most notable is the absolute decline in the amount of income from agriculture. Grain income alone dropped from 175 yuan in 1987 to 113 in 1999.[39] This 35 percent decline can be attributed primarily to the collapse in grain prices described earlier in the paper. Other sources of agricultural income – with the minor exception of income from fruit – declined to the extent that the overall share of farm income declined to 33 percent of total income, a drop of twenty percentage points from 1987. What is especially important to note is that the decline in this share is NOT due to merely increasing relative importance of non-farm income, but an absolute decline in levels of agricultural income. Moving down the column, we see improvements in income from family businesses, in absolute terms from 91 to 162 yuan, and from 16 to 23 percent as a share of total income. But the largest improvements in family income came from wage earnings, especially wages earned outside the village. The wage earnings outside the village include household members still resident in the village, but who commute outside the village to work, as well as wage earnings remitted by non-resident (but still locally registered) household members. The RCRE survey does not permit a further

---

[39]While the 1987 to 1999 trend suggests a straight decline of grain income from 175 to 113 RMB per capita, grain income peaked at 257 RMB per capita in 1995, and declined rapidly between 1995 and 1998.

disaggregation. Clearly, however, locally earned wages have become less important in both relative and absolute terms, while employment opportunities outside the village and accessed through migration have become the dominant source of labor earnings.

The Shorrocks decompositions are presented in Table 6. As a general summary, controls for location rarely matter, indicating that composition of income matters within-villages much the same way as across villages. Furthermore, the OLS and 2SLS estimates generally agree, at least in terms of broad conclusions. We indicate where the OLS and 2SLS coefficients are statistically significantly different by a *. To limit the numbers to discuss, we focus on the 2SLS results.

For 1987, we find that agricultural income, while dis-equalizing, contributed less to overall inequality than its share of total income (19 percent versus approximately 40 percent). The same applied to agricultural sidelines, so that only 21 percent of total inequality was attributed to inequality of agricultural income, even while this source accounted for 53 percent of total income. Non-farm family businesses contributed most to inequality compared to their share of income (27 percent compared to about 16 percent), followed by wage income (31.6 percent compared to 25.1 percent). Within the wage category, local wages were relatively disequalizing, while wages from employment outside the village were relatively equalizing.

The results for 1999 are significantly different, and even more different than the change in average composition would suggest. First note that inequality of agricultural income contributed only 3.5 percent of overall income inequality. Even adding livestock and other sidelines, the overall contribution of farming income to inequality was 6.3 percent. It would seem that to the extent that the machinery of redistribution (restrictions on farm size, for example) is directed towards minimizing inequality of farm income, it is misdirected. Inequality of non-farm family business income contributes more to inequality in 1999 than 1987, though this is not surprising given its increased importance as a source of income. Perhaps the most striking result of the decompositions is the large share – 47.5 percent, or almost half – of total inequality attributed to wage earnings. Local wage earnings, while they have declined in magnitude, are relatively unequally distributed, and disproportionately earned by higher income households. Inequality of wage earnings from outside the village explains 21.4 percent of overall inequality, but this is actually lower than its share of total income. To this extent, access to these wage opportunities is relatively equalizing. Note also that this is one example where controls for village dummies make some difference, as the within-village contribution (13.3 percent) is less than the total contribution, reflecting spatially uneven development of off-farm labour markets.

Taken together, these decompositions highlight two important sources of inequality, especially when we compare 1999 to 1987. First is the sharp decline of the relatively equalizing source of income from farming. Second is the relative increase in disequalizing income from non-farm family businesses,

and the failure of non-farm labour markets to provide income opportunities for low income households that offset the collapse of agricultural income. Past emphasis on the role of non-farm income as a source of inequality was only partially correct: these results suggest that given the recent trajectory of farm income, efforts to improve the rural distribution of income should be placed on improving access to non-agricultural employment for low income households. Increasing agricultural incomes – at least in an equalizing way – are unlikely to improve overall income distribution, if for no other reason than agricultural incomes are only weakly associated with overall income, and they are also very low.

5.0   Discussion and Conclusions

There is certainly a risk of over-simplification in attempting to summarize our key findings. After all, the underlying data are based on household surveys with about 8000 observations per year for twelve years, and such measures as the Gini coefficient are summaries themselves that obscure the complexities of income distributions. That said, our most striking finding is the unambiguous deterioration of the income distribution over the last half of the 1990's. Concern with rising inequality is justified, especially when combined with declining living standards for those at the lower end of the distribution.

The most obvious next question is why this deterioration has occurred: have economic reforms failed? Have market reforms created an economy that disproportionately rewards winners and heavily penalizes losers? Our results provide some insights towards these questions. First, we rule out geography as the most important factor for understanding the dispersion of incomes: at a point in time, more than half of inequality is due to inequality between neighbors within a village, not differences in income between rich and poor regions. Furthermore, we find that regional income differences are declining over time. If most inequality is within-villages, then this should turn our attention to determinants of within-village inequality, such as village-level institutions, market development, and the distribution of household endowments. An important avenue for future research is to document the joint evolution of village incomes and the distribution of village income, including a careful assessment of causal linkages between village growth and inequality, as well as other correlates of village-level growth and inequality. For example, very little is known about the role of education, and the potential interaction of human capital with market development and access to non-farm opportunities.[40] Efforts to design appropriate social safety nets and to improve local tax policy in rural areas need to be informed by a better

---

[40]Benjamin, Brandt, Glewwe, and Li (2002) show exploratory results that suggest that the combination of rising education levels and the development of non-farm employment opportunities can reduce inequality, and furthermore, that the effect of the distribution of education on overall income distribution depends on local development of markets.

understanding of the ways in which local institutions and markets influence prospects for reductions in poverty and inequality, and improve the growth prospects of the local economy.

Second (like previous researchers), we confirm that non-agricultural incomes are an important source of inequality. Indeed, to the extent that studies use NSB-like data (including the RCRE), both the level and trend of this source of inequality may be understated. But it would be a mistake to conclude that runaway income growth in non-farm income drives the winner-loser divergence in rural areas. Certainly, inequality driven by households at the very top of the income distribution is associated with lucrative family businesses. However, access to non-agricultural employment -- possibly in other people's family businesses, and in particular, employment outside the home county -- seems to be relatively equalizing. Rising inequality and falling incomes at the lower end of the distribution, are not correlated with too much non-agricultural income, but too little.

This conclusion is emphasized by our third key finding that the collapse in living standards since 1995 is driven by falling agricultural incomes. Given that output has not generally fallen, most of the decline in incomes can be attributed to low crop prices. An important area of future research thus concerns the determinants of farm prices. Are the low prices in the last half of the 1990's a transitory shock, reflecting temporary global market conditions? This may be the case, as crop prices have shown some recovery in 2003. Or are they low more permanently, because improvements in farm productivity have changed the terms of trade between agricultural and non-agricultural goods within China? If crop prices are likely to be low (though possibly fluctuating) in the near future, then this raises a number of difficult policy questions. Almost all rural income-support policies are based on guaranteeing households access to land on an approximately per capita basis, through village land allocation. While this provides households a means to feed themselves, when crop-prices are low, the value of this income support is quite low. Whatever the possible merits of this in-kind transfer for minimizing poverty, it has obvious limitations as a redistribution mechanism. With low returns in agriculture, a land policy which attempts to equalize farm incomes will have only a weak impact on overall inequality, given the small and declining share of income earned in agriculture.

The potentially perverse distributional consequences of the egalitarian land allocation institution are magnified when we account for possible efficiency losses associated with restrictions on land ownership and farm consolidation. For example, restrictions that limit the development of land rental markets may limit the ability of highly productive farmers to exploit a comparative advantage in agriculture and thus allow someone to earn a profitable living in farming. To the extent that these

households are the ones without lucrative family businesses, the ability to specialize in agriculture might improve prospects for poorer households.[41]

A related big picture question is whether continued tolerance of egalitarian land redistributions by village leaders – combined with other mobility restrictions – is slowing a longer-run transition of individuals out of agriculture, by preventing farm consolidation and providing incentives for households to stay even marginally engaged in agriculture through guaranteed access to land. An interesting research program would be to compare China's current situation with other economies' transitions from agricultural to industrial economies. For example, the percentage of the labor force in agriculture in the United States declined from 64 percent in 1850 to 38 percent in 1900, 12.2 percent in 1950, and only 2.4 percent today.[42] This process took place more slowly than China's transition, but was far from painless, and it would be interesting to see to what extent falling farm incomes (in the United States) and rising non-farm opportunities facilitated the transition. To the extent that existing policy in China is slowing down the adjustment process, it provides an opportunity to explore the interactions between an apparently benign redistribution policy (egalitarian land distribution), and the longer run evolution of income distribution and growth. It goes without saying that beyond facilitating households' participation in non-farm opportunities, more immediate poverty-alleviation mechanisms need to be considered.

Finally, our results show that before the big picture can be fully understood, there are a number of important data and measurement issues to be confronted. An important starting point would be improved access to NSB household survey data, so that richer cross-time and cross-space comparisons can be made. As good as the RCRE data are, it would be helpful to broaden participation in the evaluation of poverty and inequality policy by opening up the NSB to more users. In addition, our comparison with other surveys shows the importance of measuring income from family businesses in understanding overall inequality, especially at the top end of the distribution. Combined with insights from other (more comprehensive) studies, a richer set of questions should be included in the NSB surveys to track this important source of income.[43] Even with their current limitations, however, results based on the RCRE survey should raise some alarm at the fate of the poor in rural China, both for their own sake, as well as for the sustainability of future reforms.

---

[41]See Benjamin and Brandt (2002) for a discussion of the efficiency costs associated with the current land property rights and allocation regime, as well as potential distributional implications.
[42]See "A History of American Agriculture," by the Economics Research Service, US Department of Agriculture (http://www.usda.gov/history2/text3.htm).
[43]See Vijverberg and Mead (2000), for example.

6.0   References

Aghion, Philippe, Eve Caroli and Cecilia Garcia-Penalosa (1999). "Inequality and Economic Growth: The Perspective of the New Growth Theories," *Journal of Economic Literature* v37, n4 (December 1999): 1615-60.

Banerjee, Abhijit V. and Esther Duflo (2000). "Inequality and Growth: What Can the Data Say?" National Bureau of Economic Research Working Paper: 7793 July 2000.

Bramall, Chris (2001). "The Quality of China's Household Income Surveys," *China Quarterly* v0, n167 (September 2001): 689-705.

Brandt, Loren and Carsten Holz (2003). "Spatial Price Differences in China." University of Toronto, manuscript.

Benjamin, Dwayne and Loren Brandt (1997). "Land, Factor Markets, and Inequality in Rural China: Historical Evidence," *Explorations in Economic History* v34, n4 (October 1997): 460-94.

Benjamin, Dwayne and Loren Brandt (2002): "Property Rights, Labour Markets, and Efficiency in a Transition Economy: The Case of Rural China," *Canadian Journal of Economics*, 35(4), November, pp. 689-716.

Benjamin, Dwayne, Loren Brandt, Paul Glewwe and Guo Li (2002). "Markets, Human Capital and Income Inequality in an Economy in Transition: The Case of Rural China," in *Inequality Around the World*, Richard Freeman (ed.), Palgrave-MacMillan (2002).

Cai, Fang, Dewen Wang and Yang Du (2002). "Regional Disparity and Economic Growth in China: The Impact of Labor Market Distortions," *China Economic Review* v13, n2-3: 197-212.

Chen, Shaohua and Ravallion, Martin (1996). "Data in Transition: Assessing Rural Living Standards in Southern China" *China Economic Review* v7, n1: 23-56.

Dayton-Johnson, Jeff and Pranab Bardhan (2002). "Inequality and Conservation of the Local Commons: A Theoretical Exercise," *Economic Journal* v112, n481 (July 2002): 577-602.

Deaton, Angus (2003). "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)," NBER Working Paper, #9822.

Deaton, Angus (1997). *The analysis of household surveys: A microeconometric approach to development policy*, Johns Hopkins University Press for the World Bank, 1997.

Deininger, Klaus and Lyn Squire (1996). "New Data Set Measuring Income Inequality," *World Bank Economic Review* v10, n3 (September 1996): 565-91.

Gustafsson, Bjorn and Li Shi (2002). "Income Inequality within and across Counties in Rural China 1988 and 1995," *Journal of Development Economics* v69, n1 (October 2002): 179-204.

Haddad, Lawrence; Kanbur, Ravi (1990). "How Serious Is the Neglect of Intra-Household Inequality?" *Economic Journal* v100, n402 (September 1990): 866-81.

Hare, Denise (1994). "Rural Nonagricultural Activities and Their Impact on the Distribution of Income: Evidence from Farm Households in Southern China," *China Economic Review* v5, n1 (Spring 1994): 59-82.

Hutzler, Charles (2003). "For China's Poor Regions, a Kinder, Gentler Beijing --- Incoming Leadership Makes A Priority of Those Ignored By Headlong Liberalization," *The Wall Street Journal*, page A13, 25 February 2003.

Jalan, Jyotsna and Martin Ravallion (2002): "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China," *Journal of Applied Econometrics*, 17, no. 4, July-Aug. 2002, pp. 329-46

Jenkins, Stephen P. (2000). "Modelling Household Income Dynamics," *Journal of Population Economics* v13, n4 (December 2000): 529-67.

Khan, Azizur Rahman, Keith Griffin and Carl Riskin (1999). "Income Distribution in Urban China during the Period of Economic Reform and Globalization," *American Economic Review* v89, n2 (May 1999): 296-300.

Khan, Azizur Rahman and Riskin, Carl (1998). "Income and Inequality in China: Composition, Distribution and Growth of Household Income, 1988 to 1995," *China Quarterly* v0, n154 (June 1998): 221-53.

Kanbur, Ravi and Xiaobo Zhang (1999). "Which Regional Inequality? The Evolution of Rural-Urban and Inland-Coastal Inequality in China from 1983 to 1995," *Journal of Comparative Economics* v27, n4 (December 1999): 686-701.

Kuznets, Simon (1955). "Economic Growth and Income Inequality," *American Economic Review* 45: 1-28.

Meng, Xin (2003). "Economic Restructuring and Income Inequality in Urban China," *Review of Income and Wealth* (forthcoming).

Morduch, Jonathan and Terry Sicular (2002). "Rethinking Inequality Decomposition, with Evidence from Rural China, *Economic Journal* v112, n476 (January 2002): 93-106.

National Statistics Bureau (NSB) (2003). China Statistical Yearbook 2003 [Zhongguo tong ji nian jian 2003]. China National Statistics Bureau (formerly State Statistical Bureau), National Statistics Bureau Publishing House, Beijing.

Ravallion, Martin and Shaohua Chen (1999). "When Economic Reform Is Faster Than Statistical Reform: Measuring and Explaining Income Inequality in Rural China," *Oxford Bulletin of Economics and Statistics* v61, n1 (February 1999): 33-56.

Rawski, Thomas G. (2001). "What Is Happening to China's GDP Statistics?" *China Economic Review* v12, n4 (2001): 347-54.

Rozelle, Scott (1994). "Rural Industrialization and Increasing Inequality: Emerging Patterns in China's Reforming Economy," *Journal of Comparative Economics* v19, n3 (December 1994): 362-91.

Shorrocks, Anthony F. (1982). "Inequality Decomposition by Factor Components," *Econometrica* v50, n1: 193-211.

Shorrocks, Anthony F. (1983). "The Impact of Income Components on the Distribution of Family Incomes," *Quarterly Journal of Economics* v98, n2: 311-26.

Tsui, Kai-yuen (1998). "Trends and Inequalities of Rural Welfare in China: Evidence from Rural Households in Guangdong and Sichuan," *Journal of Comparative Economics* v26, n4 (December 1998): 783-804.

Vijverberg, Wim and Donald Mead (2000): "Household Enterprises," in Grosh, Margaret and Paul Glewwe (eds), *Designing household survey questionnaires for developing countries: Lessons from fifteen years of the Living Standards Measurement Study*. Vol 2. 2000, pp. 105-37. Washington, D.C.: World Bank.

World Bank (2003). *China: Promoting Growth with Equity, Country Economic Memorandum*, Poverty Reduction and Economic Management Unit, East Asia and Pacific Region, Report No. 24169-CHA.

Table 1
Per Capita Income and Consumption: Levels and Growth
Selected Years, RCRE

	Spatial Deflator?	1987	1991	1995	1999	Implied Growth Rate
Income	No	578	551	772	714	0.018
	Yes	557	530	750	691	0.018
Consumption	No	410	402	548	508	0.018
	Yes	396	392	537	494	0.018
Observations		7,983	7,903	6,738	6,987	

Notes: This table shows man real per capita household income and consumption (in constant 1986 RMB yuan) for selected years. The implied growth rate is defined as the average annual compound growth rate that would turn 1987 incomes to 1999 levels. The spatially deflated rows adjust for regional price differences using the price deflator in Brand and Holz (2003), and described in the data appendix.

Table 2
Per Capita Income and Consumption Inequality
Various measures and Selected years, RCRE

	1987	1991	1995	1999
Income				
Gini (NOT Spatially deflated)	0.32	0.33	0.33	0.37
Gini (Spatially Deflated)	0.29	0.30	0.31	0.35
Other measures of inequality:				
Variance of Logs	0.57	0.60	0.59	0.73
Atkinson (Sensitivity=2)	0.28	0.32	0.33	1.00
Atkinson (Sensitivity=1)	0.16	0.17	0.17	0.21
Atkinson (Sensitivity=.5)	0.08	0.09	0.09	0.12
Percent Below Half Cont Mean	0.14	0.17	0.15	0.20
Percent Below 1987 Half Cont Mean	0.14	0.19	0.07	0.14
90th/10th Split	4.06	3.98	3.93	5.24
Consumption				
Gini (NOT Spatially deflated)	0.25	0.27	0.27	0.31
Gini (Spatially Deflated)	0.22	0.24	0.25	0.28
Other measures of inequality:				
Variance of Logs	0.44	0.47	0.47	0.55
Atkinson (Sensitivity=2)	0.18	0.20	0.21	0.27
Atkinson (Sensitivity=1)	0.09	0.11	0.11	0.14
Atkinson (Sensitivity=.5)	0.05	0.06	0.06	0.08
Percent Below Half Cont Mean	0.08	0.09	0.08	0.12
Percent Below 1987 Half Cont Mean	0.08	0.10	0.02	0.06
90th/10th Split	3.09	3.22	3.28	4.06

Notes:: This table provides various distributional summary statistics corresponding to the mean per capita income and consumption levels reported in Table 1. We show (1) The Gini coefficient, repeated for spatially undeflated and spatially deflated levels; (2) The variance of log per capita income and consumption; (3) The Atkinson Index, calculated with three inequality aversion parameters (decreasing in aversion for 2.0, 1.0, and 0.5); (4) The proportion of households with incomes below one-half the contemporaneous mean income (i.e., the 50 percent of mean income that year); (5) The proportion of households below one half the mean income level for 1987 (an approximation to a constant "poverty line"); and (6) The ratio of the 90th to the 10th percentiles.

Table 3
Comparison of Surveys: Levels of Income and Inequality

Survey:	CHNS 1997		RCRE				CCAP			
			1997		1999		2000a		2000b	
	Mean	% not 0	Mean	% not 0	Mean	% not 0	Mean	% not 0	Mean	% not 0
Income, by source:										
Agriculture	816	78.7	624	96.1	507	94.8	606	90.4	607	90.9
Livestock	92	50.0	212	77.8	182	75.2	211	75.2	211	76.4
Wages	764	37.3	748	64.9	851	68.0	892	62.4	893	62.5
Other	386	51.8	148	86.0	150	80.3	155	52.6	155	34.0
Family Business	418	21.9	522	54.5	494	50.1	796	29.7	534	28.2
Total Income	2,477	97.9	2,255	100.0	2,184	99.9	2,667		2,370	
Inequality: Gini	0.43		0.36		0.37		0.50		0.44	

Notes: This table compares levels, composition, and inequality of per capita income for three data sets (surveys) at similar (though not identical) points in time. All reported values are in nominal (undeflated) terms. For each data set, we report mean values of per capita income by source, as well as the percentage of households with non-zero observations ("% not 0"). The three primary data sets are China Health and Nutrition Survey (CHNS) for 1997; (2) The RCRE for 1997 and 1999; and the Chinese Centre for Agricultural Policy (CCAP) survey for 2000. Results from the CCAP survey are shown for 2000a, a data set that includes all households with positive income, and for 2000b, based on the same sample, but excluding the top one percent of per capita income households. Further details are available in the Data Appendix.

Table 4
Contribution of location to variance of income and consumption
RCRE, Selected Years

	1987	1991	1995	1999
Dependent variable ln (Income per capita)				
without spatial deflator				
Contribution of region	0.186	0.162	0.154	0.120
Contribution of province	0.237	0.218	0.183	0.153
Contribution of village	0.500	0.466	0.413	0.424
with spatial deflator				
Contribution of region	0.082	0.070	0.074	0.054
Contribution of province	0.119	0.092	0.083	0.073
Contribution of village	0.422	0.380	0.342	0.370
Dependent variable ln (consumption per capita)				
without spatial deflator				
Contribution of region	0.190	0.184	0.162	0.181
Contribution of province	0.278	0.246	0.189	0.231
Contribution of village	0.560	0.529	0.507	0.525
with spatial deflator				
Contribution of region	0.051	0.063	0.064	0.085
Contribution of province	0.137	0.102	0.083	0.117
Contribution of village	0.474	0.439	0.442	0.454

Notes: This table shows the fraction of variation of real log per capita income (and consumption) attributed to location. This is simply the R-squared from a regression of log per capita income on a set of location dummies. The decompositions are reported with or without the income variable spatially deflated. The effect of location is reported at three levels of aggregation: (1) the village (XX villages); (2) Province (nine provinces, as described in the Data Appendix); and (3) Region, defined as West (Gansu, Shanxi, and Sichuan), Central (Anhui, Henan and Hunan) and East (Guangdong, Jiangsu and Jilin).

Table 5
The composition of income in 1987 compared to 1999
(in 1986 RMB)

|	Mean	1987 Share	% > 0	Mean	1999 Share	% > 0	Growth
Total Income	578	1.000	1.000	714	1.000	0.999	0.018
Agricultural Income	229	0.397	0.981	158	0.222	0.942	-0.031
Grain Income	175	0.303	0.978	113	0.158	0.926	-0.037
Cash Crop Income	46	0.080	0.812	30	0.042	0.564	-0.036
Fruits, Tea and Dates	8	0.014	0.248	15	0.022	0.257	0.056
Agricultural Sidelines	74	0.129	0.955	68	0.095	0.764	-0.007
Forest Products	17	0.029	0.333	10	0.014	0.185	-0.043
Livestock	54	0.093	0.950	50	0.069	0.746	-0.007
Aquaculture	4	0.007	0.125	8	0.012	0.058	0.062
Family Businesses	91	0.157	0.616	162	0.227	0.501	0.048
HH Industry	27	0.048	0.135	44	0.061	0.073	0.039
Construction	6	0.010	0.072	11	0.016	0.046	0.056
Transportation	17	0.030	0.065	26	0.037	0.076	0.036
Commerce, Service & Trade	25	0.042	0.126	57	0.079	0.172	0.070
Other Family Business Income	16	0.027	0.395	24	0.034	0.251	0.036
Wage Income	145	0.251	0.711	276	0.387	0.680	0.054
Local Wage Income	85	0.147	0.452	79	0.111	0.257	-0.006
Employment Outside	46	0.080	0.390	175	0.245	0.505	0.111
Local Gov't Employment	14	0.024	0.069	22	0.031	0.049	0.038
Family Transfers	29	0.050	0.525	34	0.048	0.495	0.016
Government Transfers	4	0.008	0.651	6	0.008	0.708	0.022
Other Income	5	0.009	0.138	8	0.012	0.110	0.040

Notes: This table compares the composition of income in 1987 to 1999. Real per capita income is shown for detailed sub-categories of income, along with the share of income ("Share") accounted for, and the proportion of households with non-zero income for that source. The last column reports the implied annual growth rate of income for that source. Note that wage income is divided between "Local" wage income, and "Employment Outside". Employment outside is employment outside of the village, and in most cases outside the township. This category includes both family members employed in distant migrant labor markets and those who are commuters returning to the village on weekends.

Table 6
Shorrocks Decompositions
RCRE, 1987 and 1999

|	1987				1999			
	(1) Share	**(2)** OLS	**(3)** 2SLS	**(4)** 2SLS	**(5)** Share	**(6)** OLS	**(7)** 2SLS	**(8)** 2SLS
Village Dummies?	N/A	NO	NO	YES	N/A	NO	NO	YES
Agricultural Income	0.397	0.126	0.190*	0.171	0.222	0.045	0.035*	0.040
Grain Income	0.303	0.050	0.090*	0.099	0.158	0.002	0.004	0.021
Cash Crop Income	0.080	0.050	0.075*	0.061	0.042	0.026	0.022	0.008
Fruits, Tea and Dates	0.014	0.026	0.026	0.011	0.022	0.017	0.009*	0.011
Agricultural Sidelines	0.129	0.074	0.106*	0.104	0.095	0.061	0.028*	0.038
Forest Products	0.029	0.009	0.025*	0.033	0.014	-0.002	-0.006*	0.007
Livestock	0.093	0.047	0.062*	0.064	0.069	0.036	0.013*	0.019
Aquaculture	0.007	0.018	0.019	0.007	0.012	0.027	0.021	0.012
Family Businesses	0.157	0.345	0.232	0.268	0.227	0.438	0.391*	0.409
HH Industry	0.048	0.140	0.092*	0.115	0.061	0.213	0.169*	0.175
Construction	0.010	-0.004	0.000	-0.002	0.016	0.015	0.011	0.002
Transportation	0.030	0.077	0.027	0.020	0.037	0.062	0.031*	0.017
Commerce, Service & Trade	0.042	0.109	0.093	0.121	0.079	0.122	0.154*	0.203
Other Family Business Income	0.027	0.023	0.020	0.015	0.034	0.026	0.026	0.012
Wage Income	0.251	0.373	0.376	0.316	0.387	0.400	0.475*	0.401
Local Wage Income	0.147	0.298	0.310	0.270	0.111	0.150	0.204	0.170
*Employment Outside**	0.080	0.062	0.044	0.009	0.245	0.206	0.214	0.133
Local Gov't Employment	0.024	0.013	0.023*	0.037	0.031	0.045	0.057*	0.098
Family Transfers	0.050	0.076	0.090	0.122	0.048	0.043	0.056*	0.094
Government Transfers	0.008	0.000	0.000	0.008	0.008	0.002	0.004	0.009
Other Income	0.009	0.007	0.005*	0.010	0.012	0.011	0.010	0.009

Notes: The table shows Shorrocks Decompositions, described in th text. Household per capita income by source is regressed on total per capita income. Columns 2 and 6 show OLS coefficients of income per capita, and columns 3 and 7 show the same coefficients, but with income per capita instrumented by consumption per capita, as a "control" for measurement error in income. Asterisks indicate where the OLS and 2SLS coefficients are significantly different (using a standard Hausman test) and thus that the 2SLS coefficients are to be preferred. Columns 4 and 8 show 2SLS results with village dummy variables added to control for the possible geographic differences of income composition. For reference, the share of income by source is reported (and is the same as in Table 5).

**Figure 1: Comparing Income Trends in the RCRE and NSB Surveys**
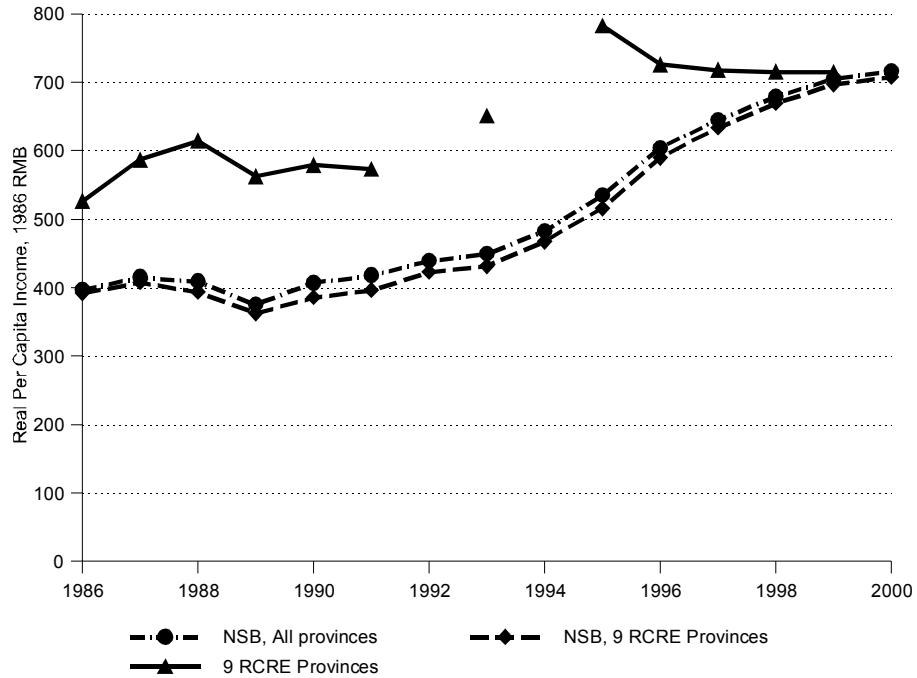
Notes: This figure compares real per capita household income levels and trends in the RCRE surveys with corresponding numbers reported in the NSB yearbooks. The NSB results are shown for all provinces, as well as the same nine provinces in the RCRE sample. All figures are deflated to 1986 RMB yuan.

**Figure 2: A Comparison of Inequality Trends based on RCRE
and NSB Households Surveys**
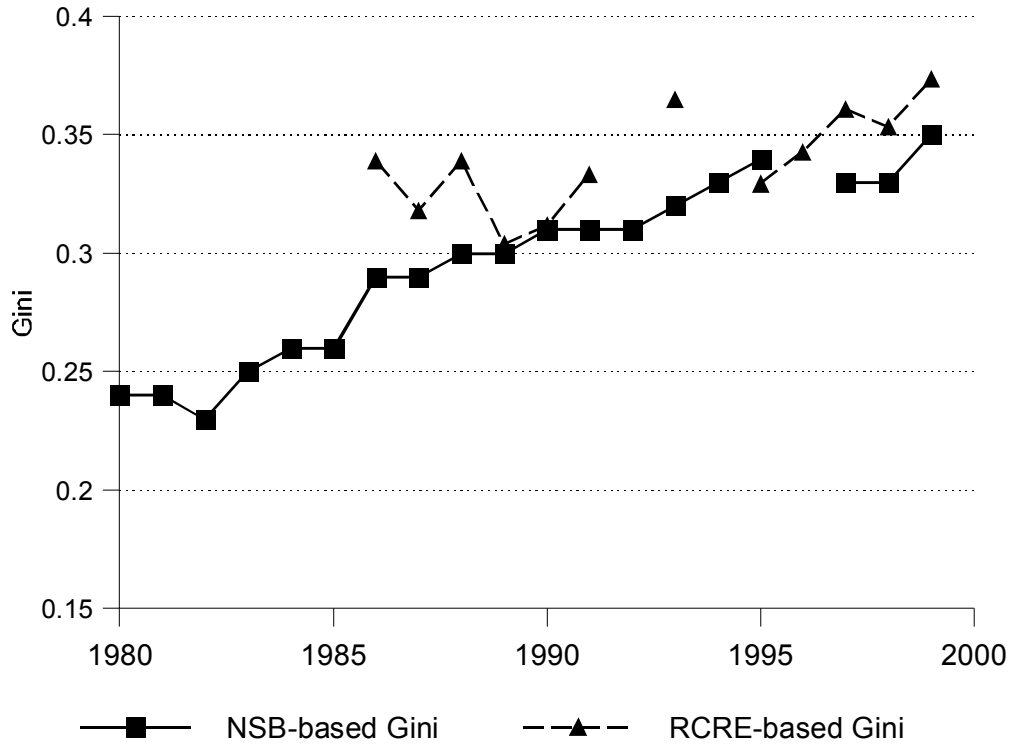
Notes: This figure compares the Gini coefficients for household per capita income that we calculated using the RCRE, with Gini coefficients based on the national sample NSB data. The NSB results come from Bramall (2001).

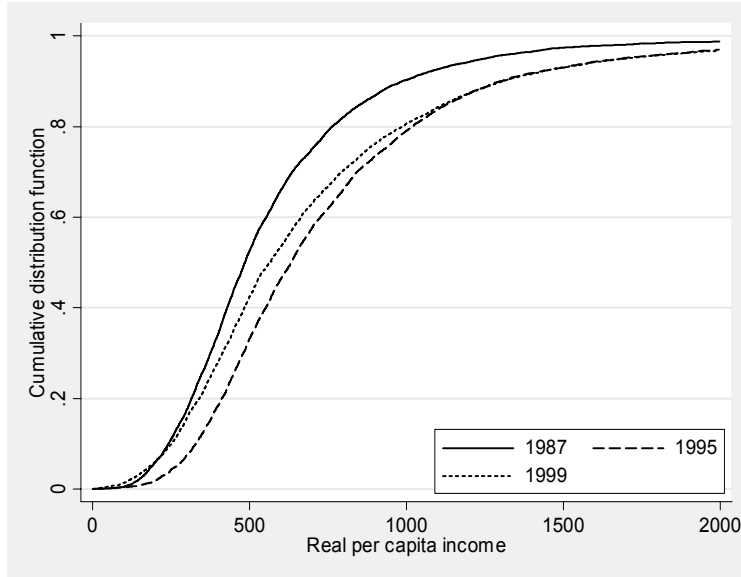**Figure 3: Lorenz Curves for Per Capita Income, Selected Years**
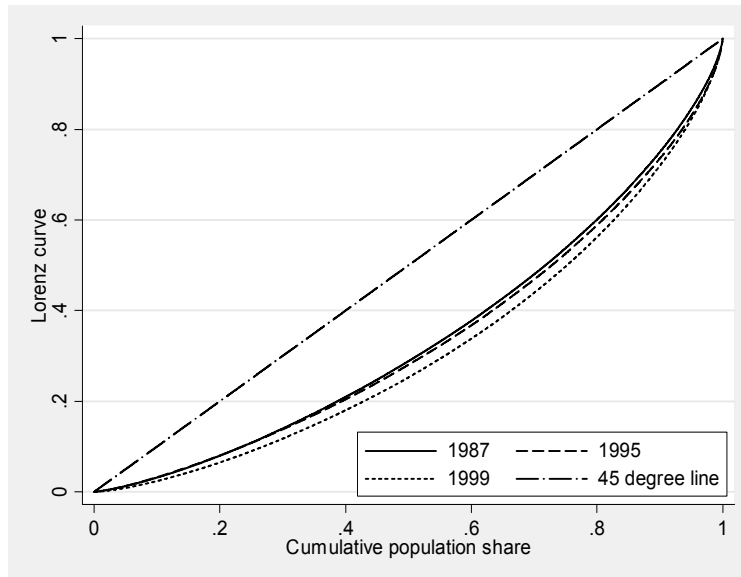
Panel A: The full picture
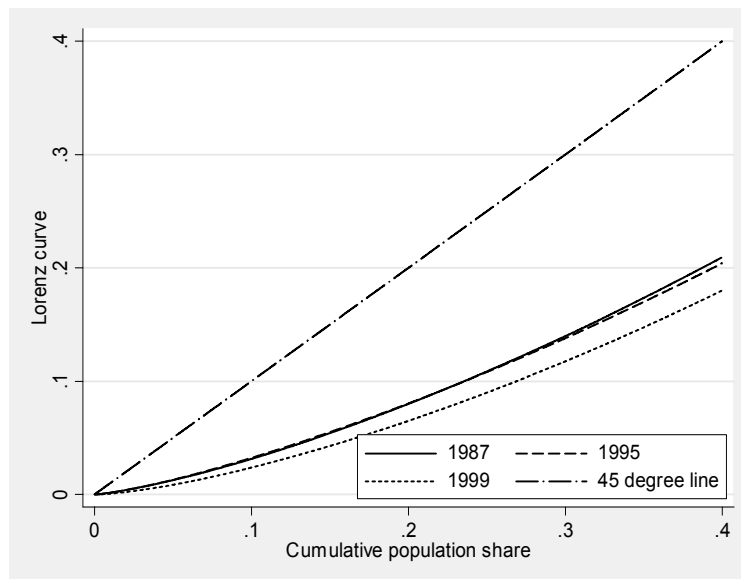
Panel B: The bottom 40 percent (from Panel A)

Notes: This figure shows the Lorenz curves for the distribution of real household per capita income for selected years, using the RCRE survey data. The top panel shows the entire curves (with the reference 45-degree line), while the bottom panel "magnifies" the curves, focusing on the poorest 40 percent of households.

Figure 4: Generalized Lorenz Curves for Per Capita Household Income, Selected Years
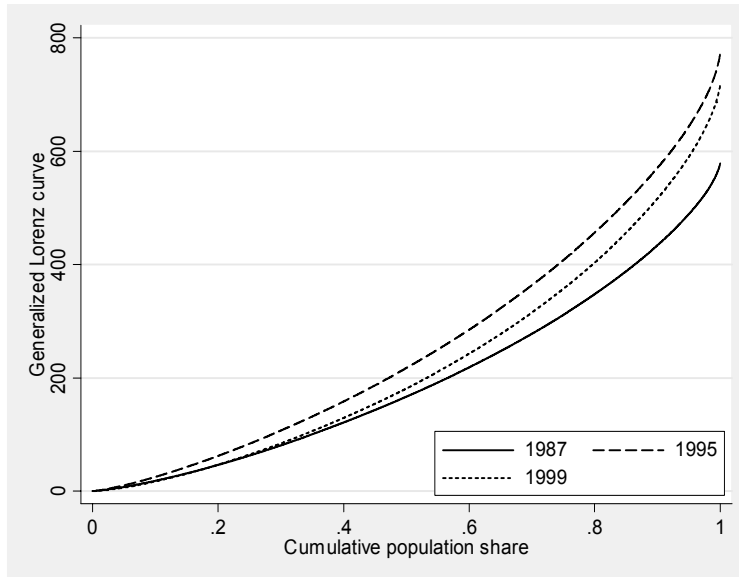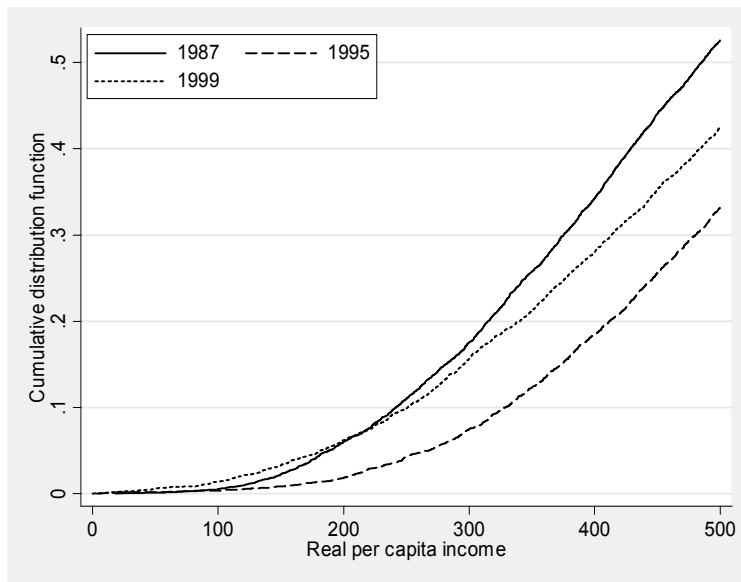

Panel A: The full picture
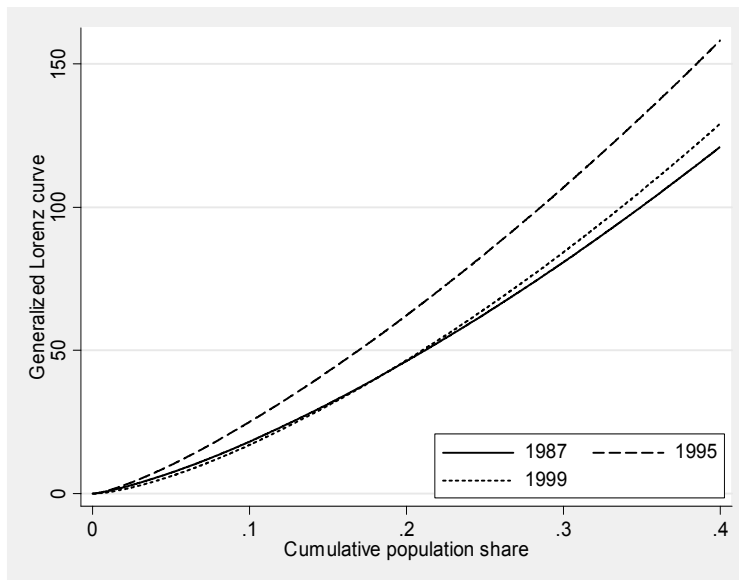
Panel B: The bottom 40 percent (from Panel A)

Notes: This figure shows the Generalized Lorenz curves for the distribution of real household per capita income for selected years, using the RCRE survey data. The Generalized Lorenz curve is simply the Lorenz curve multiplied by mean income. The top panel shows the entire curves (with the reference 45-degree line), while the bottom panel "magnifies" the curves, focusing on the poorest 40 percent of households.

Figure 5: The CDF of Per Capita Household Income, Selected Years


Panel A: The full picture


Panel B: The bottom 40 percent (based on Panel A)

Notes: This figure shows the Cumulative Distribution Function (CDF) for the distribution of real household per capita income for selected years, using the RCRE survey data. The top panel shows the entire curves, while the bottom panel "magnifies" the curves, focusing on the poorest 40 percent of households.

Figure 6: Comparing Inter-Provincial Inequality in the RCRE and NSB Surveys
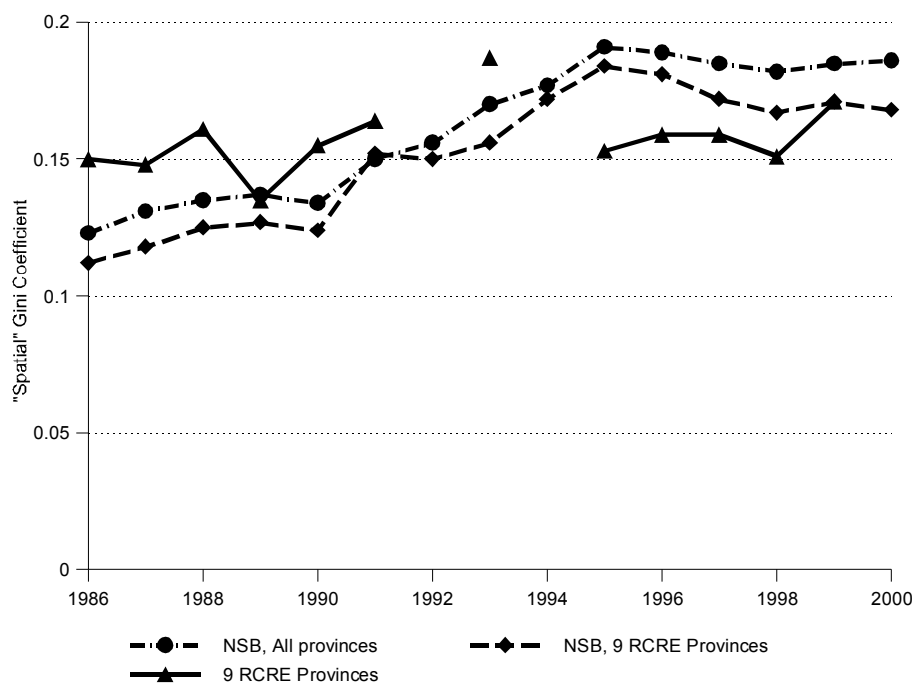

Note: This figure compares the amount of implied interprovincial inequality in the RCRE and NSB surveys. Interprovincial inequality is calculated on the basis of attributing everyone in a given province with the mean per capita income, and calculating the implied Gini with provincial population weights. This "simulation" exercise is conducted for all provinces in China using NSB-reported mean provincial incomes; for the subset of provinces covered by the RCRE sample (but using NSB mean incomes); and for the RCRE sample.

Figure 7: Growth in Per Capita Income by Initial Per Capita Income, RCRE Provinces
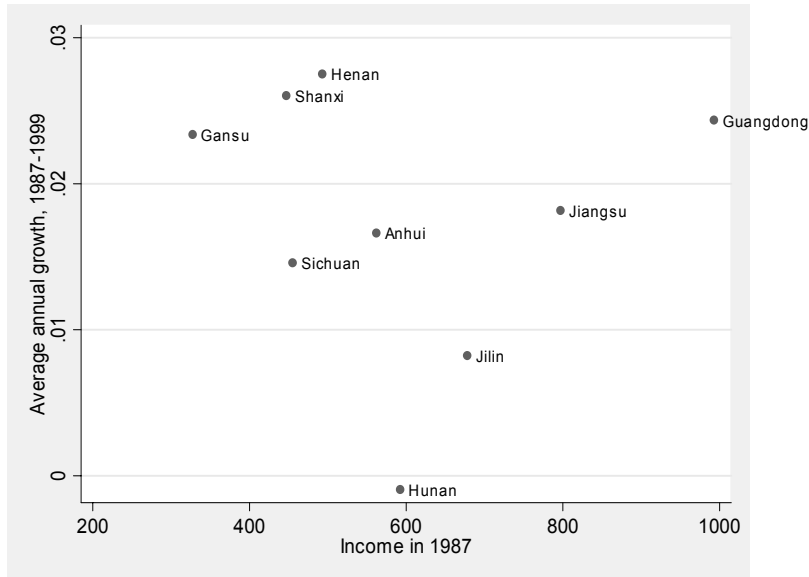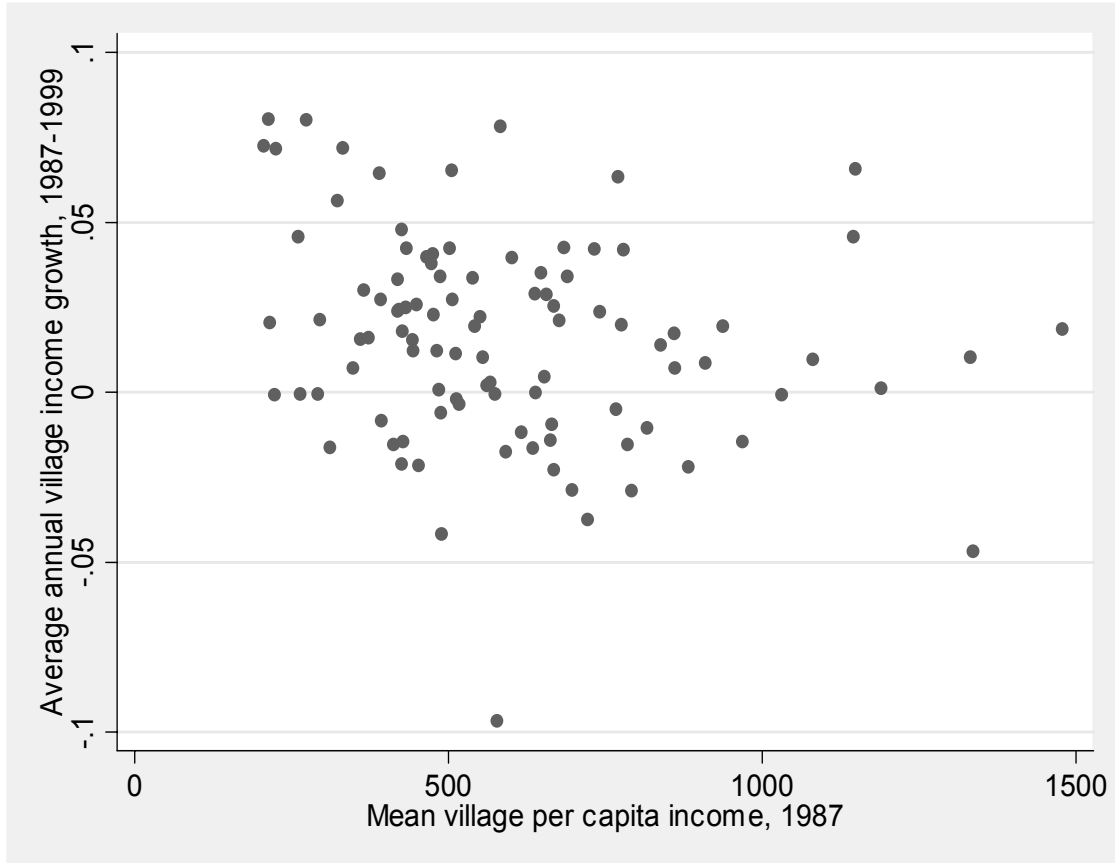

Notes: This figure arrays average annual growth rates for incomes by province (based on RCRE data) by "initial income," that is, mean provincial income in 1987.

Figure 8: Changes in Provincial Inequality, 1987 to 1999
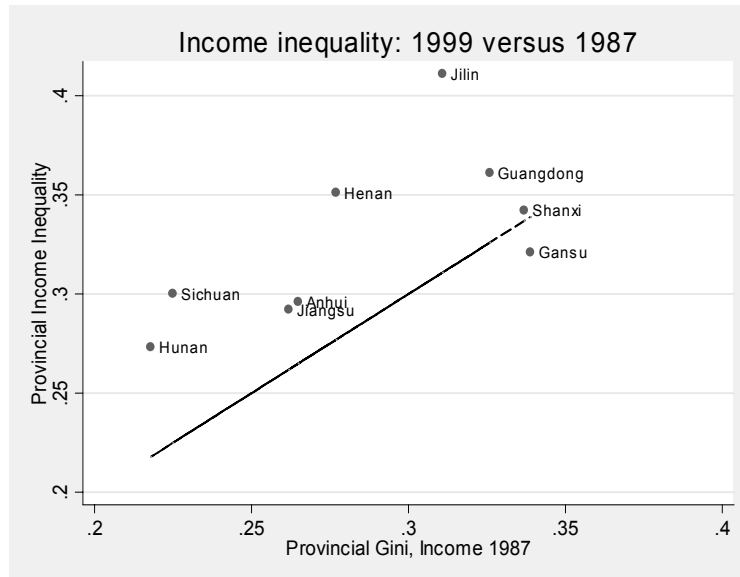

Notes: This figure plots the provincial Gini for per capita income in 1999 (calculated with the RCRE data) versus the provincial Gini for 1987. The 45-degree line serves as a reference, whereby points lying above the 45-degree line correspond to increases in provincial inequality between 1987 and 1999.

Figure 9: Within-Village Growth Versus Initial Village Income

Notes: This figure arrays average annual growth rates for incomes by village (based on RCRE data) by "initial income," that is, mean village income in 1987. The points in this figure can be used to estimate a "convergence" regression. Such a regression yields Growth = $0.14 - 0.02$ lnY87 (t=2.9).

Figure 10: Evolution of Within-Village Income Inequality
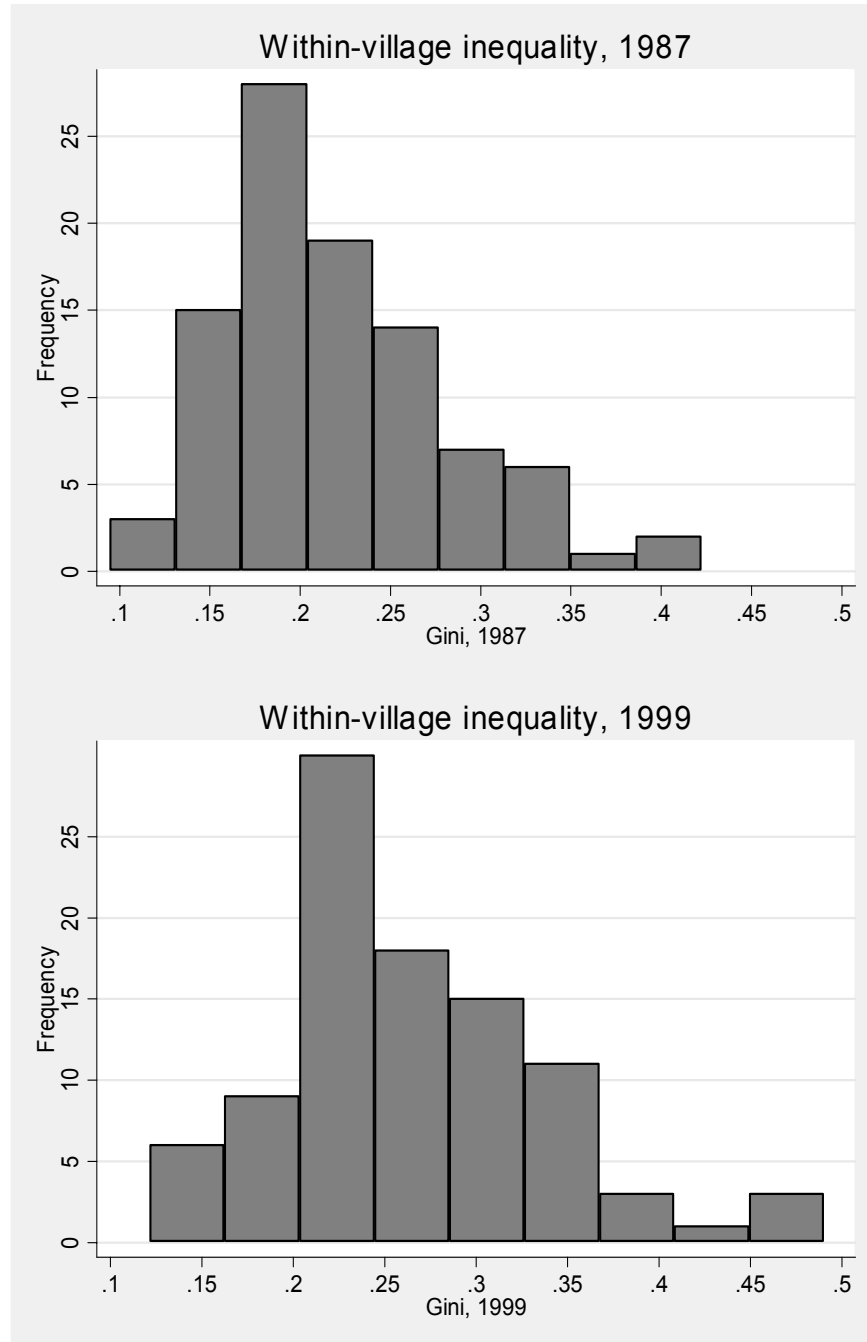
Notes: These histograms report the frequency of various magnitudes of village-level Gini coefficients, for 1987 and 1999.

Figure 11: Changes in Village Inequality Versus Initial Inequality
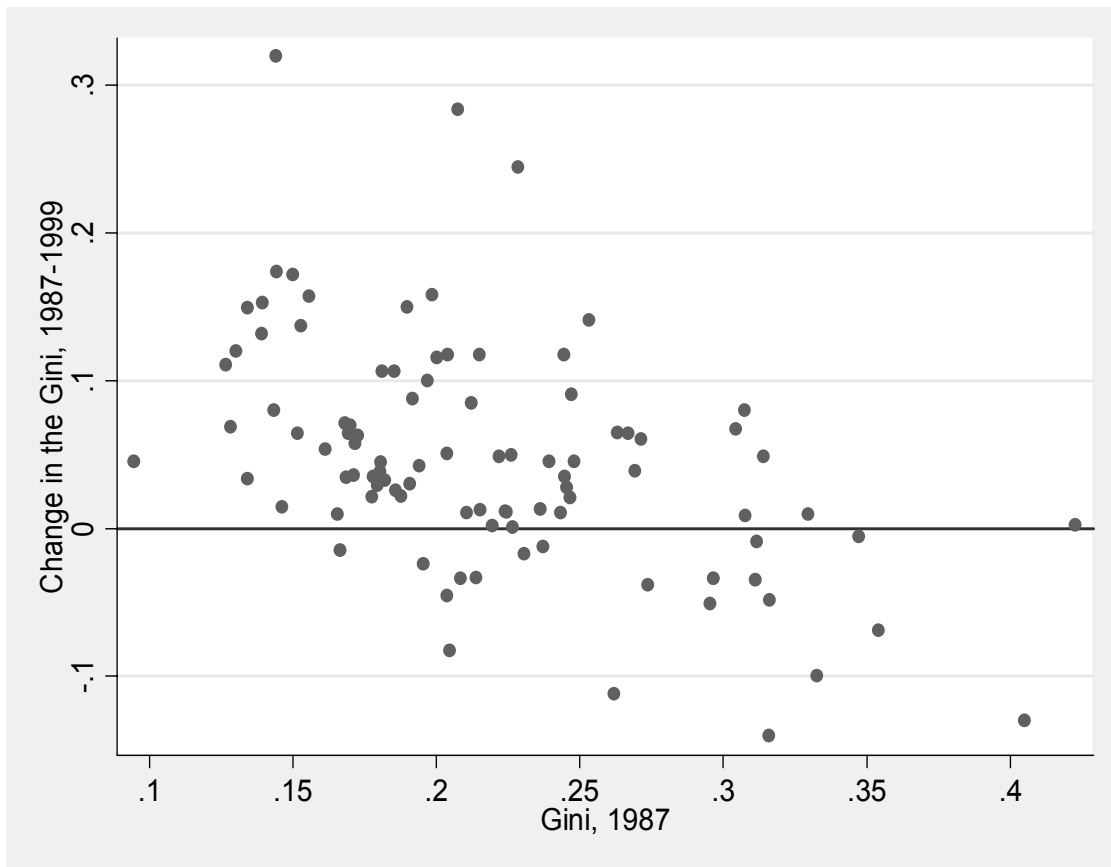
Notes: This figure plots the changes in village-level inequality between 1987 and 1999 (based on the village-level Gini) versus the initial level of inequality (the Gini in 1987). This figure illustrates "convergence" of inequality levels across villages.

7.0 Data Appendix

7.1 RCRE Village Locations: Province and Region

The data for the analyses of this paper come from nine provinces of the Research Center for Rural Economy (RCRE) village and household surveys. Basic information on sampling within province and region is provided in appendix table A.1. We follow the literature in grouping provinces into regions. On average, RCRE surveyed households in 30 villages in both the western and the eastern region, and 45 villages of the central region. Management of the survey was delegated to provincial offices, which made decisions regarding within-province sampling rates. In each province, an equal numbers of poor, medium and rich counties were selected, from each of which a village of average socioeconomic status was surveyed. At the village level, between 10 and 20 percent of households, or roughly 30-130 households, were then randomly selected.[44]

7.2 Attrition of Households and Villages from the RCRE Surveys

RCRE first fully implemented the national survey in 1986, doubling the number of villages in 1987. Since then, there has been relatively little change in the number of sampled villages. Attrition of villages from the survey has occurred, however, primarily for two reasons. First, RCRE's mandate is to use the survey to study agricultural production and factors influencing changes in agricultural productivity. Over the period from 1986 to 1999 four villages in Jiangsu and two in Guangdong were dropped and replaced because they were no longer engaged in agriculture. Second, attrition has also occurred as a result of disagreements between county or village leaders and provincial administrators of the survey. Of the 103 villages in the survey at the end of 1999, 82 have been in the survey since 1987. A significant amount of attrition of villages occurred during gaps when the survey was not conducted in 1992 and 1994. Much less change in villages (and households) occurred during the periods without gaps from 1987 to 1991 and 1995 to 1999, for which 98 and 97 villages, respectively, were in the sample for all five years. In principle, dropped villages were to be replaced by a representative village in the same county of the same average per capita income.

Attrition has also occurred at the household-level (a detailed table is available upon request), and averages roughly five percent per year. Considerably more attrition came during the two-year gaps, and is largely associated with the loss of entire villages. Our estimate of attrition is also conservatively high. Households with the same household identifier in two successive years, but with significant differences in demographic structure, characteristics of housing or economic activities, were treated as separate households. In these cases, we treat the year *t* household as a new observation and consider the year *t-1* household to have dropped from the sample. For the entire period between 1987-1999, we have a full

---

[44]Our sample originally included Zhejiang province. An examination of county gross value of output and mean rural per capita income revealed that nearly 90 percent of the surveyed counties in the province were from the upper-third of the distribution. Thus, they were unlikely to represent well the distribution of income within Zhejiang, leading to biased estimates of inequality across regions. Sampling of counties in other provinces appears to be consistent with RCRE's guidelines.

panel of 4352 households. For the two sub-periods, namely, 1987-1991 and 1995-1999, panel size is 6691 and 5796 households, respectively

7.3 The Sample Used in Our Analyses

In the sample we use, we trim extreme outliers from the dataset because we suspect coding errors or errors in which fixed investment is inappropriately coded as an operating cost in the household budget. To identify potential outliers, we first calculate median income and consumption per capita in each village for each year. We then drop households if the absolute value of the difference between household reported income per capita and village median income per capita for the year is greater than five times village median income. We apply the same criteria to household consumption per capita. In each year, less than one tenth of one percent of households were dropped under these criteria; altogether, 382 observations were dropped over the 1987-1999 period.

Our "Full trimmed" sample is thus this trimmed household sample, with all available observations, including panel, attrited, and replacement households. We also performed the analysis with the panel households only ("Trimmed Panel"), which would be most sensitive to possible biases introduced by attrition, and the results are similar.

7.4 Issues in the Calculation of Household Income and Consumption per Capita

Grain crops remain an important component of household production, yet in the RCRE survey grain produced for own consumption or stored is valued at prices reflecting the quota price rather than the market price. Up through the mid-1990s, quota prices were well below market prices. For this reason, income from grain production and consumption out of home production are both likely biased downward. To deal with this problem, we re-value the household's non-marketed grain (and grain consumption out of own production) at average village market prices.

7.4.1 Definition of Income and Consumption Per Capita

Household income is the sum of income from all household-managed activities (farming, agricultural sidelines, and non-agricultural activities), local wage employment, migrant remittances, formal transfers from the village and subsidies from higher levels of government, and informal transfers from friends or family (but excluding borrowing). Consumption is calculated as the sum of expenditures on food and non-durable goods purchased during the year, the value of home-produced goods consumed, the value of the flow of services from the household's stock of durable goods and housing, and the value of services (education, health care and other) purchased by the household during the period. Nominal values are converted into 1986 RMB using the provincial rural CPI from National Statistical Bureau yearbooks.

7.4.2 Consumption (Durables and Housing)

Our measure of household consumption per capita includes the value of the flow of services from the stock of consumer durables and housing. The RCRE surveys provide estimates of the original value of housing and durable goods, and report current expenditures on durables and new investment in expanding houses. To value the flow of services from housing and durables, we must first use this information to come up with a reasonable estimate of the current value of housing and durables, and then estimate the flow value of consumption.

<u>Durable Goods</u>. We assume that durable goods (and production assets) were accumulated in equal portions over the years between 1978 and the first year that the household appeared in the survey. We assume further that durable goods and production assets have a useful life of seven years (we checked

robustness using five and ten year lives), and that the non-depreciated portion of the durable good maintains its "real" value. For each year we depreciated one-seventh of the current value of the good, appreciated the remaining value of the good using a rural provincial capital goods price index, and added the new durables accumulated during that year. From this annual value of the stock of durables, we assume that the household consumes one-seventh of the existing stock of durable goods during the current year.

Housing. We assume that housing is consumed over a twenty-year period. For the initial year of the survey and for the first year that a new household appears in the survey, we value the housing stock using information from other households in the village on the real cost per square meter of living area in new housing constructed in the village in year *t-1, t* and *t+1*, and the livable floor space of the household. For each succeeding year we subtract one-twentieth of the estimated value of the house (as depreciation), appreciate the remaining 19/20[ths] of the value of the house using the rural capital goods price index, and add on the real value of new additions to the house made during the year. The current flow consumption of housing is one-twentieth of this current year value of housing.

Discussion of CHNS and CCAP Surveys and Sample Sizes in Table 4

The CCAP rural household survey was carried out in the six provinces of Liaoning, Hebei, Shanxi, Sichuan, Hubei and Zhejiang and covers the year 2000. Altogether, 1200 households in 60 villages were surveyed, or 20 households per village. In each province, counties were stratified on the basis of the gross value of agricultural and industrial output, and one county selected from each of the five quintiles. Within each county, townships were similarly stratified, and a township selected from both the upper and bottom-half of the distribution. A village was then randomly selected from each of the townships. Within each village, households were randomly selected on the basis of the most recent village household registry. Household membership was defined in a manner analogous to that used by the RCRE survey.

Details on the CHNS survey can be found on the CHNS website, http://www.cpc.unc.edu/projects/china/

Spatial Deflator

In order to control for absolute differences in price levels facing households across provinces, we deflate rural incomes and consumption using a spatial deflator constructed by Brandt and Holz (2003) for 1990. Using the NSB rural household survey expenditure data, they construct a single nationwide consumption basket that includes food, clothing, articles daily use, energy, services, housing, and durable goods. The basket is then priced using provincial-level price data. The range across the nine provinces in 1990 in the absolute price level is more than fifty percent (Guangdong, 1.38, and Sichuan, 0.88).

Table A.1
Basic Statistics on Sample by Province and Region

Region/Province	1987	1989	1991	Year 1993	1995	1997	1999
Western Region							
Number of Villages	32	34	34	34	31	30	30
Avg HHs/Village	424.2	452.6	525.0	471.2	493.5	503.4	508.9
Avg Sampled							
HHs/Village	56.1	56.4	55.7	53.0	55.7	55.3	55.1
Central Region							
Number of Villages	48	48	49	44	44	44	44
Avg HHs/Village	313.4	332.0	353.2	367.0	405.3	408.2	403.6
Avg Sampled							
HHs/Village	72.2	72.1	72.5	63.3	63.7	63.9	63.4
Eastern Region							
Number of Villages	31	30	30	29	26	31	29
Avg HHs/Village	455.7	451.9	475.7	508.0	457.3	461.1	503.4
Avg Sampled							
HHs/Village	87.6	80.6	82.1	72.7	74.8	70.8	75.2
Overall Total							
Number of Villages	111	112	113	107	101	105	103
Avg HHs/Village	385.1	400.7	437.4	438.3	445.7	451.0	462.0
Avg Sampled							
HHs/Village	71.9	69.6	70.0	62.5	64.1	63.5	64.2

Province-by-province annual information on attrition are available by request.

DAVIDSON INSTITUTE WORKING PAPER SERIES - Most Recent Papers

The entire Working Paper Series may be downloaded free of charge at: www.wdi.bus.umich.edu

CURRENT AS OF 2/26/04

Publication	Authors	Date
No. 654: The Evolution of Income Inequality in Rural China	Dwayne Benjamin, Loren Brandt and John Giles	Feb. 2004
No. 653: The Sources of Real Exchange Fluctuations in Developing Countries : an Econometric Investigation	Imed Drine and Christophe Rault	Feb. 2004
No. 652: Ownership, Control and Corporate Performance After Large-Scale Privatization	Jan Hanousek, Evzen Kocenda and Jan Svejnar	Feb. 2004
No. 651: Czech Social Reform/Non-reform: Routes, Actors and Problems	Jiri Vecernik	Feb. 2004
No. 650: Financial Integration, Exchange Rate Regimes in CEECs, and Joining the EMU: Just Do It...	Maurel Mathilde	Feb. 2004
No. 649: Corporate Investments, Liquidity and Bank Financing: Empirical Evidence from an Emerging Market	Arun Khanna	Feb. 2004
No. 648: Financial Constraints in Investment – Foreign Versus Domestic Firms. Panel Data Results From Estonia, 1995-1999.	Tomasz Mickiewicz, Kate Bishop and Urmas Varblane	Feb. 2004
No. 647: Legal Minimum Wages and the Wages of Formal and Informal Sector Workers in Costa Rica	T. H. Gindling and Katherine Terrell	Feb. 2004
No. 646: Employment Expectations and Gross Flows by Type of Work Contract	Catalina Amuedo-Dorantes and Miguel Á. Malo	Jan. 2004
No. 645: Transition on the Shop Floor - The Restructuring of a Weaving Mill, Hungary 1988-97	János Köllő	Jan. 2004
No. 644: Fighting "Low Equilibria" by Doubling the Minimum Wage ? Hungary's Experiment	Gábor Kertei and János Köllő	Jan. 2004
No. 643: Mind the Break! Accounting for Changing Patterns of Growth during Transition	Jan Fidrmuc and Ariane Tichit	Jan. 2004
No. 642: The Monetary Approach to Exchange Rates in the CEECs	Jesús Crespo-Cuaresma, Jarko Fidrmuc and Ronald MacDonald	Jan. 2004
No. 641: Community Norms and Organizational Practices: The Legitimization of Wage Arrears in Russia, 1992-1999	John S. Earle, Andrew Spicer and Klara Sabirianova Peter	Jan. 2004
No. 640: Trust in Transition: Cross Country and Firm Evidence	Martin Raiser, Alan Rousso, and Franklin Steves	Jan. 2004
No. 639: What Makes Small Firms Grow? Finance, Human Capital, Technical Assistance, and the Business Environment in Romania	J. David Brown, John S. Earle and Dana Lup	Jan. 2004
No. 638: The Politics of Economic Reform in Thailand: Crisis and Compromise	Allen Hicken	Jan. 2004
No. 637: How Much Restructuring did the Transition Countries Experience? Evidence from Quality of their Exports	Yener Kandogan	Jan. 2004
No. 636: Estimating the Size and Growth of Unrecorded Economic Activity in Transition Countries: A Re-Evaluation of Eclectric Consumption Method Estimates and their Implications	Edgar L. Feige and Ivana Urban	Dec. 2003
No. 635: Measuring the Value Added by Money	Vlad Ivanenko	Nov. 2003
No. 634: Sensitivity of the Exporting Economy on the External Shocks: Evidence from Slovene Firms	Janez Prašnikar, Velimir Bole, Aleš Ahcan and Matjaž Koman	Nov. 2003
No. 633: Reputation Flows: Contractual Disputes and the Channels for Inter-firm Communication	William Pyle	Nov. 2003
No. 632: The Politics of Development Policy and Development Policy Reform in New Order Indonesia	Michael T. Rock	Nov. 2003
No. 631: The Reorientation of Transition Countries' Exports: Changes in Quantity, Quality and Variety	Yener Kandogan	Nov. 2003
No. 630: Inequality of Outcomes and Inequality of Opportunities in Brazil	François Bourguignon, Francisco H.G. Ferreira and Marta Menéndez	Nov. 2003
No. 629: Job Search Behavior of Unemployed in Russia	Natalia Smirnova	Nov. 2003