The Evolution of Income Distribution and Poverty in Rural China during Reform: An Empirical Evaluation

A revised proposal submitted to the PEP Research Network

August 28, 2003

Sangui Wang
Chinese Academy of Agricultural Sciences
1.0 Motivation

Over the course of the past two decades, the Chinese economy has enjoyed rates of per-capita income growth of nearly six percent. While this growth has succeeded in pulling millions of individuals out of poverty, especially in the countryside, inequality has also significantly increased, nationally, as well as separately in urban and rural areas. With the ongoing restructuring of the state-enterprise sector in the cities, the continued rapid development of China’s coastal regions, and the sharp drop in farm prices leading up to China’s entry into WTO, there are new concerns that the gap between China’s richest and poorest citizens may be widening at an even more rapid rate.

A careful and accurate description of trends in inequality and poverty levels is absolutely essential to good policy-making in China, as is an analysis of the factors underlying these trends. Moreover, with China representing nearly one-fifth of the world’s population, an accurate assessment of inequality in China is also important to related work on changes in global inequality. Yet, in estimating inequality and poverty levels, our welfare measures are only as good as the underlying survey data we have on household incomes and consumption, and the analysts’ awareness of the strengths and weaknesses of these data.¹

For a developing economy in transition such as China’s, these issues are inherently difficult. As examples: (1) for rural households in China, a significant (albeit declining) portion of activity is for own-consumption, and thus is not explicitly valued. Output gets sold at both planned (quota) as well as market prices; (2) new sources of incomes, including off-farm wage income, and that from family-run businesses, have also appeared, with survey design often lagging in its ability to capture accurately these new income; (3) differences in the rate at which

¹ See the review article by Angus Deaton (2003) for an up-to-date summary of these issues.
markets are liberalized also imply potentially important differences in price levels (and thus real incomes) across localities.

Drawing on the three major household-level surveys carried out in rural China over the last two decades, two of which we have extensive experience in working with, our purpose in this project is two-fold:

1) Provide a careful assessment of trends in inequality and poverty levels;
2) Use this as a basis for analyzing key factors associated with inequality.

The three surveys include the rural household survey carried out by the Research Centre for Rural Economy (RCRE) under the Ministry of Agriculture almost every years since 1986; the China Health and Nutrition Survey (CHNS), which has been carried out in 1989, 1991, 1993, 1997 and 2000; and the National Statistical Bureau’s (NSB) annual rural household survey. In addition to these three sources that cover several years, I also plan to employ at least two single-year cross-sections of high quality household survey that have been employed in the past. I explain their value below.

The proposal itself is divided into three key parts. First, I provide a brief summary of what we know about inequality levels and trends, and discuss several important conceptual issues relating to the measurement of inequality and poverty levels. I highlight the results of an ongoing project that forms the foundation of the research we propose here. I next provide the rationale for using the NSB rural survey data in conjunction with several other sources of rural survey data. The key innovation of the research we propose is the incorporation of the NSB data, and a significant part of the budget of the project is devoted to the costs of obtaining access to these data. And third, I discuss in more detail the kind of analysis we intend to carry out as we explore the factors underlying inequality trends.
2.0 What do we know about inequality in China?

We know comparatively little about the evolution and nature of income inequality in China, even by developing country standards. To a large extent, this is a consequence of restrictions of access to nationally representative household survey data (Deaton, 2003). As a result, the frontier of knowledge concerning inequality in China is relatively uneven, and usually based on small, region-specific household surveys covering a few, or even one year. Clearly, results based on these surveys may not be nationally representative, and surveys covering a short time spans or only a couple of years may poorly reflect trends. Single-year surveys from different years may suffer from a host of comparability problems.

The NSB collects rural household surveys covering 60,000 households every year, is nationally representative, and while it has evolved over time, is conducted in a manner that maximizes comparability across years. While they do not make these data generally accessible, they do publish aggregate summary statistics, for example, means and Gini coefficients for per capita household income. In Table 1 I show these NSB-published Gini coefficients, which provide first-order evidence of inequality trends in rural China for the past 15 years. As the table shows, inequality has risen from 0.29 to 0.35 since 1987. These numbers allow us to conclude that inequality “has gone up”, but little else.

Other researchers have used different data sets or subsets of the NSB data to explore deeper the nature of inequality over this time period. Chen and Ravallion (1996) for example, have used the NSB data for four southwestern provinces covering the period between 1985 and 1990. Wang and Wen (2003) used data from the China Poverty Survey to analyze changes between 1997 and 2000, while Kahn and Riskin (1998) used NSB survey data from a subset of provinces for 1988 and 1995, augmented with additional household information. Gustafsson and
Li (2002) use the same data to document differences in inequality between 1988 and 1995. As did Benjamin, Brandt, Glewwe and Guo (2002) using the CHNS, Gustafsson and Li (2002) point out that the conventional wisdom that inequality is driven by regional income differences is probably wrong (more on this point later). Benjamin, Brandt, Glewwe and Guo (2002) also highlight the important role of non-farm income (as do others in the literature), with a detailed evaluation and comparison of measurement issues in the CHNS and other surveys. Despite differences in data sets and methodology, all of these studies concur that inequality has “gone up.” They also raise other questions/problems concerning the role of geography and non-farm income. One objective of our current research is to address these questions with a common time series of data sets.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini for per capita income, NSB national sample ¹/</td>
<td>0.29</td>
<td>0.31</td>
<td>0.34</td>
<td>0.35</td>
</tr>
<tr>
<td>Gini for per capita income ²/</td>
<td>0.32</td>
<td>0.33</td>
<td>0.33</td>
<td>0.37</td>
</tr>
<tr>
<td>Gini for per capita income (spatially deflated) ³/</td>
<td>0.29</td>
<td>0.30</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>Percentage with income below one-half 1987 mean income ⁴/</td>
<td>0.14</td>
<td>0.19</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>Gini for per capita consumption ⁵/</td>
<td>0.25</td>
<td>0.27</td>
<td>0.27</td>
<td>0.31</td>
</tr>
<tr>
<td>Gini for per capita consumption (spatially deflated) ³/</td>
<td>0.22</td>
<td>0.24</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>Percentage with cons. below one-half 1987 mean cons. ⁴/</td>
<td>0.08</td>
<td>0.10</td>
<td>0.02</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Notes:
1/ Source: Bramall (2001).
2/ Based on full sample, reported in Benjamin, Brandt, and Giles (2003).
3/ Spatial deflator based on Brandt and Holz (in progress).
4/ Defined as 0.5 * mean income in 1987 (in real terms).
5/ Consumption includes current expenditures plus the imputed value of services from durables and housing.

Ideally, such a study would use a time-series of cross-sections based on raw household-survey data covering every year from the beginning of reforms to the present. No such “ideal” studies exist, but to my knowledge, the closest approximation (in terms of data) is a recent paper by Benjamin, Brandt, and Giles (2003, in progress) that employs household survey data from the RCRE.
In their paper, they construct measures of household income and consumption (as alternative measures of household financial well-being) in order to evaluate the robustness of conclusions concerning trends in inequality. A selection of their results is reported in Table 1, and their conclusions basically mirror those of the aggregate NSB. However, this over-summarizes the issues that need to be considered in drawing such a conclusion. More precisely (to mention a few):

- **Valuation of home consumption:** A considerable (though declining) portion of household production is consumed by the farmer and his/her family. In the early reform period, the NSB valued this production at quota (administered) prices, whereas a more appropriate value would be the cost to the farmer of purchasing the food (grain) in the market. This issue was addressed by Ravallion and Chen (1996 and 1999) in their use of NSB subsets of household data, and by Benjamin, Brandt and Giles (2003) in their use of the RCRE. The issue is most important for early years (and thus the base-line for the trend), and in particular, for the poorer households that consume a higher fraction of their farm production.

- **Income versus consumption:** Which is the best measure of living standards that we are most concerned about “inequality”? By some criteria, consumption is less volatile than income, and more tied to long-run income potential, and should be the focus of analysis. Furthermore, households are more prone to understating their income than consumption, so consumption measures may be more reliable. On the other hand – especially if personal savings are an important source for investment into future income – income may better reflect long run earnings potential. In addition, by looking at the sub-components of income, we can better understand why inequality and average incomes are changing. On the basis of the RCRE, as
can be seen in Table 1, the consumption-based Gini’s are lower than the income-based Gini’s (as usual), but the trend is the same.

• Inequality versus poverty: If inequality is rising, but everyone is getting richer, the consequences are quite different than when some individuals are being left behind with incomes below an absolute minimum standard. The careful measurement of poverty requires the construction of poverty lines (and the underlying “poverty basket”), an income level below which we define an individual as poor. While we plan on constructing such lines for the NSB data (drawing considerably on my own past research\(^2\)), a simple exercise shows the importance of doing so. I define a “line” equal to one-half the mean income in 1987. In 1987, 14 percent of individuals had incomes below this level. As incomes rose over the 1990’s, we can track how overall living standards might rise despite the increase in inequality. In 1991, the percentage below the “line” increased (with the recession), though by 1995 the fraction of “poor” declined to half the rate of 1987. However, by 1999, the poverty rate returned to 14%. This implies that poverty and inequality both deteriorated over the last half of the 1990s, which is a shocking result. It also highlights the importance of going beneath the aggregate statistics provided by the NSB, as well as the need to evaluate the sensitivity of conclusions concerning the poor to assumptions about how their income is measured (i.e., the valuation of home consumption).

• Spatial and temporal deflation: 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

\(^2\) See the references listed below for examples of Wang’s experience in estimating Poverty trends for rural China: (Park and Wang (2001); Park, Wang, and Guobao (2002); Wang (2000); Wang, Park, and Guobao (1999); Wang and Wen (2003))
inequality can be biased. For example, if price levels and incomes are positively correlated so that prices are higher in richer areas, failure to account for price differences will lead to an overestimate of inequality. Chen and Ravallion (1995) and Tsui (1998) both construct spatial deflators for their analysis, albeit, for only four and two provinces, respectively.

- **Housing and durables**: A potentially important component of a household’s income is the annualized value of owner-occupied housing services. These services, along with that derived from other household durables, make up an integral part of household consumption, especially for higher income households. In their recent work with the RCRE data, Benjamin, Brandt, and Giles constructed estimates of these sources of income and consumption. The value of these services has not been formally incorporated into the NSB estimates, but the raw data collected by NSB will allow their estimation.

- **Measurement error**: A significant component of inequality may be driven by measurement error. In fact, there is always some concern that dispersion of incomes is driven mostly by measurement error. Perhaps this explains why income-based inequality is higher than consumption-based inequality. That would suggest that our estimate of income inequality is too high. However, our experience is that the measurement error is more complicated, and generally leads to an understatement of inequality. In their evaluation of the RCRE data, for example, Benjamin, Brandt, and Giles compare income levels and inequality to other household surveys, like CCAP 2000 (collected by Loren Brandt and other researchers affiliated with the Chinese Centre for Agricultural Policy (CCAP) in 2000). They discovered that surveys like the RCRE (and NSB) that ask individuals a single (or limited set) question on how much they earned in a family business systematically understate earnings from this source. This could derive from deliberate shading of income by respondents, or just as likely,
flaws in the survey, reflecting the need for sophisticated survey instruments that collect detailed accounting information on family-based rural enterprises (as done by CCAP). In a side-by-side comparison, they found that more sophisticated surveys yield higher estimates of income inequality, for example, a Gini of 0.50. However, this is driven by 1 or 2 percent of the very rich households, and almost entirely because of an underestimate of non-farm business income. “Trimmed” estimates of the income-Gini (which exclude the top 2% of incomes, all in excess of anyone’s income in the RCRE) are closer to 0.42, and otherwise line up with the RCRE estimates of income composition. This suggests that the NSB and RCRE-based measures of inequality are too low, and underlines the value of using several data sets to corroborate conclusions. Quite simply, the RCRE and NSB data have their strengths (comparability over time), while one-time only (but intensive) surveys have theirs (richer questionnaires). Following Benjamin, Brandt, and Giles, we plan on exploiting both types of data.

Of course, our research objective goes beyond the construction of a summary table like Table 1 (difficult enough though it is to create): we wish to dig deeper in order to understand more about the nature and causes of inequality or poverty. Having household data like the RCRE allows some simple but informative exercises including spatial decompositions of income, as well as decompositions by source of income. I discuss both of these in more detail below.

As valuable as the RCRE data are, they have important limitations. For example, virtually no information is collected on the individual household members (age, sex, education, or work activity). Other data sets, while available for a single year (or two) can be used to look at cross-section relationships, such as estimating linkages between education and poverty, living standards of the elderly, or interpretations of the gap between income and consumption (the role
of savings). Benjamin, Brandt, and Giles (for example) successfully used the CHNS, and other surveys (like the CCAP 2000) to conduct this type of analysis. While these isolated surveys shed light on the trends reflected in the RCRE, obviously, it would be better to have access to the NSB which would allow an integrated cross-section and time series investigation.

3.0 Further Justification for using the NSB data

I have already come to an agreement with NSB regarding use of their rural household survey data. I have obtained full access to the raw data, however, all work must be conducted with a dedicated computer onsite at the NSB in Beijing. Given the costs involved in acquiring this access (a key motivation for our proposal), it is important to highlight the returns to working with the NSB data.

1. The NSB surveys (and summary statistics) are the data most frequently used by policymakers. It thus makes sense to use the data in which people place the most stock. Although there are certainly shortcomings with these data (like the valuation of income-in-kind), we will be able to “fix” many of these problems as has been done with the RCRE, and explain the biases introduced by neglect of these issues (and others we may not be able to solve). Moreover, in combination with the other rural households surveys, we will be able to identify other biases relating to sampling that are otherwise hard to detect (also, as we have done with the RCRE).

2. We will have access to a representative dataset for all sample households in six provinces collected by NBS, i.e. it includes provinces in every major region of China. Because the RCRE was intended as a longitudinal survey (i.e., repeated sampling of the same households), attrition over the years has possibly compromised some of it representativeness.
One benefit of using the NSB surveys is that we can evaluate the attrition problem in the RCRE.

3. In addition to the household-level data from the NSB, we will also have access to individual-level information such as age, education and occupation, that are important to any measurement and assessment of welfare, or in the estimation of household level regressions on the correlates of poverty. The RCRE data do not have this information.

4. An essential input to our analysis will be the construction of poverty lines. These require us to define a poverty basket and then to value the basket at local prices. Local price data, however, are scarce. For essential commodities such as grains, the household level data provide consumption data on expenditure and quantity consumed. These can be used to calculate unit values (expenditure divided by quantity), which can then be used to construct estimates of prices for valuing the basket. The same unit values will also be useful in constructing spatial deflators.

5. We have already obtained access to (and paid for) NSB urban data for a nationally representative sample that covers 6 major provinces plus Chongqing. (Note that the data are managed by different parts of the NSB, and different rules have been applied). With rural data from the same localities, we will also look at issues relating to overall inequality including the important issue of urban-rural differences. None of the existing research has been able to tackle the urban-rural dimension (though some comparisons can be made with the CHNS).

4.0 What we intend to do with the data

Our analysis will begin with a careful descriptive analysis of consumption and income levels, inequality and poverty levels since the mid-1980s, mirroring our analysis with the RCRE
Critical to this will be the construction of consistent estimates of real income and consumption using the NSB rural household data. Problems in the way that in-kind consumption (and its income counterpart) is valued in earlier years require us to reconstruct existing NSB consumption and income estimates. As part of this exercise, and drawing on some of the earlier work of Chen and Ravallion (1995, 1999), and Brandt and Holz (2003, in progress), we will also construct new provincial-level poverty lines and spatial deflators. The latter captures differences in the price level across space, which may distort our assessment of inequality. The former will enable us to estimate the percentage of the rural population living below the poverty line, while the latter will enable us to obtain more accurate estimates of real income and consumption. We will also experiment with alternative ways of handling income from owner-occupied housing, and consumption services from durable goods. A critical dimension to this descriptive exercise will be reconciling estimates based on NSB data with those from other surveys, like the RCRE. As noted above, we have already carried out a related exercise using the RCRE data, the CHNS data, and the CCAP 2000.

Second, we will analyze factors underlying the emerging income differences. A key part of this analysis will be decompositions of income by the source and location of income. The latter dimension allows us to determine the relative importance of income differences between versus within regions. Much of the literature concerning inequality in China has emphasized growing regional disparities as a source of the rising inequality. Yet, recent results from similar exercises using the RCRE (and other) data suggest that the role of regional (or county or village) income differences is significantly less important than commonly believed. Moreover, the role of location may have actually declined over time. Analyzing income by source, on the other hand, enables us to assess whether particular sources of income are more or less equalizing (or
These include income from farming, farm sidelines, family-run businesses, as well as off-farm wage income and remittances. The two decompositions (across space and by source of income) can also be easily combined. Information on the changes in the structure of income in rural China, combined with the results from the decomposition exercises, will provide key insight into those factors underlying the trends in inequality. For example, results from the RCRE suggest that the combination of rising non-farm income (highlighted by previous researchers) with collapsing farm incomes (from depressed grain prices) is an important “cause” of the current combination of rising inequality and poverty.

Third, we will examine the attributes of the rural poor. Drawing on the household-level data, we will examine the association between poverty and region, ethnicity, gender, age, and educational levels. Links between poverty and income sources can also be made. In this regard, (as just noted) we are especially interested in the role that declining crop prices have made to an increase in poverty levels since the mid-to-late 1990s.

5.0 Research Output

I have a specific outlet in mind for this research: An overview chapter on inequality and poverty in a forthcoming volume on China’s experience with transition, edited by Loren Brandt and Thomas Rawski. This volume (and three related conferences) has been funded by the National Science Foundation and the Smith-Richardson Foundation, both in the US. The project itself brings together leading experts inside and outside of China. We expect the book to be widely read, and the chapter on inequality to represent the “state of the art” for our understanding of the inequality and poverty experience of China through transition. Translation of the volume into Chinese is already planned.
However, the project will yield at least one further journal publication, possibly in a more general economics journal (like *Economic Development and Cultural Change* or the *Journal of Development Economics*), which has published work on inequality previously, though without the quality of data we will have. In the meantime, we will publish the working paper versions of these two articles in the widely-read William Davidson Institute working paper series, as well as the University of Toronto working paper series. Simultaneously, we will also look for leading Chinese outlets for our work.

### 6.0 Timetable

A first draft of the overview paper on inequality is to be presented in November of 2004 at the second of the three conferences planned as part of the Brandt-Rawski project. Once the funding to obtain access to the NSB rural data is secured, my research assistants and I will begin immediately working with NSB on the preliminary analysis. Our co-investigators from Canada will also be visiting Beijing: Loren Brandt has planned to China for December (2003) and February and May (2004), and Dwayne Benjamin for May (2004), both for working on the collaboration. I will also be visiting Toronto in the spring of 2004.

### 7.0 Budget

Our total budget is $20,000 (Canadian), comprised of:

1. Data access from the NSB: $15,000 (Canadian)
2. Airfare plus local costs for 2 week visit by Wang to Toronto: $2500
3. Support for research assistant in China: $2500

### 8.0 The research team
The research team is comprised of three researchers, based at the Chinese Academy of Agricultural Sciences in Beijing:

- Sangui Wang (Principal investigator)
- Ms. Li Yingxing (PhD Candidate, CAAS)
- Ms. Li Yun (Research Assistant, CAAS)

Sangui Wang is a Professor and Division Director at the Institute of Agricultural Economics, Chinese Academy of Agricultural Sciences. He is also the Director of China Poverty Research Association. Dr. Wang has done research in the filed of poverty reduction since 1988 and worked with international and bilateral organizations such as WB, ADB, UNDP, FAO, IFAD, JICA, DFID and Ausaid as poverty expert.

Ms. Li Yingxing is currently a PhD Candidate and has worked in rural finance and rural poverty for more than five years. She has considerable experiences in field research and data processing. Ms. Li Yun is a research assistant and has involved in a number of research projects in recent years. Both of them will work closely with foreign collaborators and NBS people and help with data processing and analysis.

I will also be consulting and collaborating with the co-authors of the Brandt-Rawski book chapter:

- Loren Brandt, University of Toronto
- Dwayne Benjamin, University of Toronto

While not formal members of the research team, they have considerable experience working on issues of poverty and inequality in rural China, using both historical and contemporary data. I have also collaborated with Brandt in the past, in the collection of original rural household data the last ten years. More detailed information can be obtained from their respective CVs.
Throughout the project, I plan to draw on their expertise, especially in helping our graduate students produce research that is aimed at a broader, world-wide academic and policy audience. They will be spending considerable time in China working with myself and the graduate students directly. I fully expect that this experience will contribute significantly to the professional development of the graduate students. None of my budget request pertains to their work, as they already have funding from other sources.

9.0 References


