

When Economic Reform is Faster than Statistical Reform: Measuring and Explaining Income Inequality in Rural China

Martin Ravallion and Shaohua Chen¹

Abstract

Official tabulations from household survey data suggest rising income inequality in post-reform rural China, and this trend has been of public concern. However, the structural changes in the Chinese rural economy had not been properly reflected in the methods used for processing the raw survey data. Using micro data for four provinces, we find that two-thirds of the conventionally measured increase in inequality 1985-90 vanishes when market-based valuation methods are used and allowances are made for regional cost-of-living differences. The data revisions also suggest somewhat different explanations for rising inequality. Non-farm income was secondary to grain production. Higher returns to farm land over time were inequality-increasing while lower returns to physical capital reduced inequality, as did private transfers.

¹ The authors are with the Development Research Group of the World Bank. This paper was prepared as a background paper for “China 2020: Sharing Rising Income” produced by the World Bank’s East Asia region. The authors are grateful to the report’s manager, Tamar Maluelyan, for encouraging them to investigate these issues, and for her comments on this paper. Their thanks also go to the Bulletin’s two referees for their helpful comments. The cooperation of China’s State Statistical Bureau has been invaluable. These are the views of the authors, and should not be attributed to the World Bank.

1 Introduction

There is a widespread view that the transition from a socialist economic system to a market economy will entail rising inequality, and there is support for that view in recent compilations of distributional data for the 1980s and '90s (Milanovic, 1996; Ravallion and Chen, 1997). However, these compilations are typically based on the tabulations of distributional data (drawn from household surveys) that are made available by official sources. While economic reforms often have important implications for the methods used in measuring economic welfare and inequality, government statistical agencies may not be adjusting as rapidly as one would like to the structural changes going on in the economy. And users of the official data rarely probe into the raw micro data underlying the distributional comparisons being made, either because of lack of access to the data or lack of resources for doing so.

Could lags in reforming statistical methods entail substantial biases in assessments of how inequality is changing during the transition? The structural changes going on are not necessarily inequality-increasing. A common element of socialist economic planning was the suppression of food-staple prices, to help finance industrialization.² Through market liberalization, the transition typically entails higher food staple prices. To the extent that food-staple producers are concentrated among the poor, the transition will put downward pressure on inequality. If all incomes were derived from market exchange then these effects should be seen quickly in official data on distribution drawing on household surveys. However, a large share of income in poor rural economies takes the form of direct consumption of own production. Valuations must be imputed

² This was often referred to as the “price scissors” and there is a large literature on the practice; for a recent analysis and references see Sah and Stiglitz (1992).

for this and other income sources which were not acquired through exchange. When prices are controlled by administrative fiat, the same prices are naturally used for valuation. But there can be no assurance that old administrative prices will be replaced by market prices as the transition proceeds. Unless statistical agencies are quick enough to adapt to such changes, biases can enter survey-based analyses of (among other things) income inequality.

The transition can have many other implications for measurement. The level of prices may rise faster in some regions of the economy than others after reforms (reflecting nontraded goods, or less than perfect spatial market integration, due for instance to poorly developed transportation). If it were the initially better-off regions which saw higher growth and higher inflation (due to higher aggregate demand locally) then assessments of income distribution which ignored geographic differences in prices could overestimate the rate at which inequality was increasing.

There is no good *a priori* reason to assume that there will be a bias, or that (when there is) it could go only one way. For example, the share of income from undervalued components may be no different between the rich and poor, or the rate of inflation may be higher in poorer regions. These are empirical questions, although they can be difficult to answer since they require access to, and reprocessing of, the raw data underlying official tabulations.

This paper addresses these concerns in the context of post-reform rural China. Beginning with Premier Deng's reforms in 1978, China's rural economy became market-oriented; prices were freed and the farm-household replaced the commune as the decision-making unit. These reforms brought about changes to data collection, including greater reliance on household surveys. The scope and collection methods of such surveys improved significantly during the 1980s, starting with the Rural Household Survey (RHS) introduced in 1984. This has been the main source of data for

distributional analysis on rural China.

Tabulations of results from the RHS in China's Statistical Yearbooks have suggested rising income inequality since the mid-1980s. This has been widely reported and attracted much attention.³ However, there are reasons to be cautious in interpreting the available evidence on income inequality in rural China. A number of potential problems have been identified in recent literature, including the undervaluation of income in kind from the consumption of own-farm products due to continuing reliance on planning prices for valuation purposes.⁴

³ See, for example, the front page article in *The New York Times*, December 27, 1995.

⁴ Discussions of the problems can be found in World Bank (1992), Khan et al. (1993) and Chen and Ravallion (1996).

We examine how the problems in official tabulations from the household survey data have affected measurements of the overall level of inequality, and how it changes over time. We also examine how these data problems impinge on explanations of the observed changes in overall inequality.⁵ Suppose, for example, that we want to know if the rising income inequality in China is due to the booming rural non-farm sector (including the famous Township and Village Enterprises). Or we may want to see what role public and private transfers played. In principle, the answers to such questions will depend on the method used to measure incomes at the household level. For example, undervaluing income in kind from own production might lead one to underestimate the contribution of this income component to rising income inequality, given that its progressive undervaluation over time would probably lead one to conclude (incorrectly) that this income component is becoming less covariate with total income. It is an empirical question just how robust explanations of rising inequality are to these data problems.

We address these issues using a large household-level data set for rural China spanning the period 1985-90. The region we study embraces booming Guangdong on the coast (the province surrounding Hong Kong) and the far less prosperous, and more economically stagnant, inland provinces of Guangxi, Yunnan and Guizhou. Having access to the micro data means that we can attempt to correct the main concerns about existing distributional data. After making corrections to the processing, we are able to use the survey to address a number of questions about the

⁵ There have been a number of studies attempting to throw light on the causes of inequality in China since reforms began in the late 1970s. Decompositions have been done along various dimensions (geographic and by income source) and at various levels of spatial aggregation (some by county, some by village, some household) and for differing time dimensions (some using single cross-sectional surveys, some including comparisons over time). Contributions include Knight and Song (1993), Rozelle (1994) and Howes and Hussain (1994).

proximate causes of the observed changes in income inequality.

The following section summarizes the theoretical results we will be using from the literature on inequality measurement. Section 3 then looks at the theoretical implications of undervaluing an income component for measures of inequality and their decomposition. In section 4 we describe our data, while section 5 gives our overall results on income inequality, with and without our corrections to the data processing. We then turn in section 6 to the task of explaining the observed changes in inequality. Our conclusions are summarized in section 7.

2 Inequality measurement and decomposition methods

$$I = I(y_1 / \mathbf{m}, y_2 / \mathbf{m}, \dots, y_N / \mathbf{m})$$

A measure of inequality can be written in generic form:

where y_i is the i 'th person's income in a population of size N , and \mathbf{m} is mean income. We assume that this measure is continuous, symmetric (swapping incomes does not change the measure), normalized such that inequality is zero when all persons have the same income, and that the measure satisfies the "transfer axiom" such that a transfer from rich to poor reduces inequality. For some sorts of distributional comparisons we may not need to know any more about the measure of inequality. For example, if the Lorenz curve (giving, on the vertical axis, the share of total income held by the poorest $x\%$ of the population) for distribution A is everywhere above that of B then all inequality measures in the above class of measures will show higher inequality in B than A (Atkinson, 1970).

In our empirical work we will focus on two special cases of the above class of measures.

The first is the well-known Gini index (G), given by the (household-size weighted) mean absolute deviation between all pairs of per capita household incomes. The second is a member of the Generalized Entropy class of additively decomposable measures, namely the average log deviation

$$LD = \frac{1}{N} \sum_{i=1}^N \log(\mathbf{m} / y_i)$$

of incomes from their mean:⁶

We will also be interested in explaining inequality and its changes over time. There are potentially many ways of decomposing a change in inequality by income source. Here we follow a strand of the theoretical literature which has constrained the choices by postulating certain axioms that are deemed desirable for any decomposition. (We only summarize the basic results that will be needed for the empirical work later.)

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

Let total income (per person) be divided into m categories, such that, for the i 'th household: If these components were uncorrelated with each other, and one measured inequality by the squared coefficient of variation (CV), then the natural decomposition would be to measure the contribution of each income component to inequality by its squared CV . However, in practice

⁶ If N stood instead for the number of households then household-size weights would appear in this formula. All statistics in this paper which are based on the household-level data have been household-size weighted.

different income sources are correlated to varying extents. And there are many other inequality measures that one might want to consider besides squared *CV*. How then should one apportion total inequality between components?

A powerful result proved by Shorrocks (1982) shows that a modified version of the squared-*CV* decomposition (allowing for non-zero correlations) can also be defended as a decomposition method for a wide range of inequality measures. For the class of inequality measures described above,⁷ Shorrocks shows that the proportion of total inequality contributed by

$$c_k = \frac{\text{cov}(y_k, y)}{\text{var}(y)} = \frac{r_k s_k}{s}$$

the k 'th income source is given by:

where r_k is the correlation coefficient with total income and s_k and s are the standard deviation of the k 'th income component and of total income respectively. Note that c_k sums to one over all k and is simply the ordinary least squares regression coefficient of y_k on y . The decomposition based on (4) is independent of the precise measure of inequality used (within the aforementioned class of measures).

Notice that the contribution of any income component to total inequality depends on both the variance of that component (relative to the variance in total income) and its correlation

⁷ In fact Shorrocks proves the following result for an even larger class of measures; see his paper for full details.

coefficient with total income. So the fact that some income component contributes a lot to total inequality does not necessarily mean that it is itself very unequally distributed; it may instead be highly correlated with total income, yet quite equally distributed. Similarly, a highly unequally distributed income component may contribute little to total inequality because it is roughly uncorrelated with total income, or it may be inequality-reducing because of a negative correlation with total income.

The above result holds for a decomposition of the level of inequality. What about changes in inequality over time? Building on the Shorrocks' decomposition, we follow Jenkins (1995) and Fields (1996) in calculating the contribution of the k 'th income source to the change in total

$$\mathbf{g} \equiv \frac{c_{k2} I_2 - c_{k1} I_1}{I_2 - I_1}$$

inequality between dates 1 and 2 by:

which sums to one. Notice that (unlike the levels decomposition) this decomposition will depend on the specific inequality measure used. We will compare results for the Gini index with those for the average log deviation given by (2).

One can also ask how much of the level of inequality or its change over time is due to some variable determining income through a stochastic process. To do so, replace equation (3) with a

$$y_i = \sum_{k=1}^m \mathbf{b}_k x_{ik}$$

regression model for income:

where x_{ik} is the k 'th asset (x_{im} can be taken to be an error term, with $\beta_m=1$). Following Fields

$$c_k = \frac{\mathbf{b}_k \text{cov}(x_k, y)}{\text{var}(y)}$$

(1996), the contribution of the k 'th explanatory variable to total inequality is given by:

This is simply the product of the partial regression coefficient of income on schooling (holding all other variables constant) with that total regression coefficient of schooling on income (holding nothing else constant). The contributions of each asset to the changes over time can then be determined using equation (5). The precise decomposition will naturally depend on the regression specification in (6). This should be borne in mind when interpreting the results.

3 Effects of valuation errors on measured inequality and its decomposition

It is known that inequality measures can be highly sensitive to measurement errors; a few bad observations can have a large impact on measured inequality (Cowbell and Victoria-Fester, 1996). Here we are concerned with a particular structure of measurement error, arising from undervaluation of an income component, as discussed in the introduction. We cannot find a treatment of this case in the literature, so we offer some observations, to help interpret the empirical results later. We examine effects of undervaluation on the level of inequality, the factor decomposition of inequality, and on the decomposition of changes in inequality over time.

Let us first consider the effect of the valuation error on the level of measured inequality, as this is the easiest case. The revaluation can be thought of as a negative income tax. Let the average rate of revaluation (analogous to the average tax rate) be defined as the increase in imputed

value as a proportion of original income. Following results from the literature on tax progressivity (see, for example, Pfingsten, 1988), the correction for undervaluation will lead to lower (higher) measured inequality if the average rate of revaluation falls (rises) as income increases.

What about the effect on the factor decomposition of inequality? Recall that the share of inequality attributed to a given income component is the regression coefficient of that component on total income (equation 4). Both the regressor and regressand are underestimated (by the same amount). There will be two sources of bias in the regression coefficient; the first is the usual attenuation bias due to miss-measurement of the regressor, while the second is the bias due to the fact that the same error contaminates the regressand. These two biases will work in opposite directions and so one cannot say on a priori grounds what effect this will have on the regression coefficient. Intuitively, the lower the regression coefficient, the less important will be the second source of bias. So one expects undervaluation to lead to underestimation of the contribution to inequality when that contribution is sufficiently low.

We can derive a very simple sufficient condition for signing the effect when the k 'th income

$$y_k^* = (1 + \mathbf{a}) y_k$$

component is undervalued by a constant proportion, such that the revaluation yields:

for $\mathbf{a} > 0$. We assume that $1 > c_k > 0$, although this can be relaxed; the following result holds for

$1 + 1/\mathbf{a} > c_k > -(1 + \mathbf{a}^2 v_k)/(2\mathbf{a})$ where $v_k \equiv \text{var}(y_k)/\text{var}(y)$. On revaluing the undervalued

component, its contribution to total inequality becomes:

$$c_k^* = \frac{\text{cov}(y_k^*, y^*)}{\text{var}(y^*)} = \frac{(1 + \mathbf{a})(c_k + \mathbf{a} v_k)}{1 + \mathbf{a}^2 v_k + 2\mathbf{a} c_k}$$

$$v_k > \frac{(2c_k - 1)c_k}{1 + \mathbf{a}(1 - c_k)}$$

From (9) it is readily verified that $c_k^* > c_k$ if and only if

So a sufficient condition for the undervaluation to underestimate the contribution to inequality is that the undervalued component of income accounts for less than one half of inequality.

The effect of undervaluation on a factor's contribution to changes in inequality over time ($\underline{\mathbf{g}}_k$ given by equation 5) is more complicated, since it will clearly also depend on how the factor decomposition evolves. We confine attention to the case of empirical interest later in which inequality is increasing (with or without revaluation) and the undervalued income component's contribution to inequality is underestimated. Let $\underline{\mathbf{g}}_k^*$ denote the contribution of the k 'th income

$$\underline{\mathbf{g}}_k - \underline{\mathbf{g}}_k^* = \frac{[(c_{k1} - c_{k1}^*)I_1 + (c_{k1}^* - c_{k2})I_2]I_1^* + [(c_{k2}^* - c_{k1})I_1 + (c_{k2} - c_{k2}^*)I_2]I_2^*}{(I_2 - I_1)(I_2^* - I_1^*)}$$

component to rising inequality. It is readily verified that:

If the factor decomposition does not change over time ($c_{k2}^* = c_{k1}^*$ and $c_{k2} = c_{k1}$) then clearly

$\underline{\mathbf{g}}_k - \underline{\mathbf{g}}_k^* = c_{k1} - c_{k1}^* = c_{k2} - c_{k2}^* < 0$; the undervaluation of the k 'th income component also leads to an underestimation of its contribution to rising inequality. However, the outcome is ambiguous when the factor decomposition is changing over time. From (11), the sign of $\underline{\mathbf{g}}_k - \underline{\mathbf{g}}_k^*$ will also

depend on the “cross-terms”, $c_{k1}^* - c_{k2}$ and $c_{k2}^* - c_{k1}$, at least one of which must be positive.⁸ A

$$\frac{c_{k2}^* - c_{k1}}{c_{k2}^* - c_{k2}} < \frac{I_2}{I_1} < \frac{c_{k1}^* - c_{k1}}{-c_{k1}^* - c_{k2}}$$

sufficient condition for $g_k^* > g_k$ is that:

However, it is entirely possible for revaluation to diminish the contribution of the undervalued income component to rising inequality, even when revaluation yields higher inequality at any one date. Suppose, for example, that with its undervaluation the measured contribution to inequality of the k 'th income component does not change over time ($c_{k2} = c_{k1}$), but with the revaluation its contribution is found to fall over time ($c_{k2}^* < c_{k1}^*$). Then $g_k^* > g_k$ if and only if $I_2^*/I_1^* > (c_{k1}^* - c_{k1})/(c_{k2}^* - c_{k2})$.

4 Data

The data are the household-level data from the Rural Household Survey (RHS) done by China's State Statistics Bureau (SSB). Our sample covers 9,500 rural households in Guangxi, Yunnan, Guizhou and Guangdong. The survey and steps we have taken in data processing are described in detail in Chen and Ravallion (1996). The RHS is a high quality survey in many respects, including both sampling methods and the care taken to minimize nonsampling errors

⁸ The cross terms cannot both be negative, for then [click here to view equation.](#) ↪ a contradiction.

through close supervision and regular visits to the sampled households. There are, however, problems in the methods used in processing the data after its collection, leading up to the tabulations found in the Statistical Yearbook for China. Chen and Ravallion (1996) review the main concerns about these data. We attempt to resolve the main problems by reprocessing the primary data for 1985-90.

An important concern about the official data is that they continued to rely on old planning prices for the valuation of income-in-kind from consumption of own-farm production. These prices were below market prices (and also below government procurement prices). This undervalued a large component of income ↪ notably non-marketed home production of grain ↪ and at a rising rate over time (Chen and Ravallion, 1996). The standard definitions indicate that, for our data set, an average of 21% of income came from grain production, of which 80% was the imputed value of consumption from own production. Other components of farm income appear also to have been undervalued, but this is a less worrying since the shares of income involved are smaller (22% of income came from non-grain farm output, but only 10% of this was from own consumption). Another problem is that the incomes used in past work have not included imputed rents for housing and durables. Past work has also ignored spatial differences in the cost of living.

To deal with these problems, we have revalued grain-income in kind at median local (county-level) selling prices for grain, as determined from the primary household-level data.⁹ The administrative prices conventionally used by SSB for valuation were 72% of median selling price in 1985, and this had fallen to 48% by 1990. We have also imputed rents for housing and consumer

⁹ Similar data are unavailable for revaluing other components of income in kind from own-farm production, although (as noted above) these appear to be minor.

durables, based on the asset valuations available in the primary survey data; we used five percent of the recorded dwelling value for housing and 10 percent for durables (Chen and Ravallion, 1996). And we have constructed new province-level spatial and inter-temporal cost of living indices. The spatial cost of living adjustment is based on poverty lines aiming to measure the local cost in 1988 of the same standard of living everywhere, based on a common bundle of foods and an allowance for non-food spending consistent with spending behavior at the food poverty line. The inter-temporal price indices are based on the rural CPI, though we have changed the weights to accord with consumption behavior of the poorest 30% of the population. Full details on both the poverty lines and the intertemporal cost-of-living deflators can be found in Chen and Ravallion (1996).¹⁰

To assess the contribution of these data adjustments, we give results for each of the three income definitions: The first is SSB's "net income" measure direct from the data files. We call this "original income." The second incorporates our revaluation of grain-income from own production, and imputed rents. The third uses our new deflator as well.

We use household income per person. This does not allow for economies of scale in household consumption. It is often argued that scale economies are small in low-income countries, because the share of income devoted to collectively consumed goods within the household tends to be small, although this assumption is questioned by Lanjouw and Ravallion (1995). We will consider the implications for some of our main results of allowing for scale economies, and examine how this is affected by the other data revisions.

¹⁰ A remaining limitation of these price indices is that the same deflators are used for all income groups in a given province and year. Depending on how much budget shares vary by income level, and how much relative prices change over time, this limitation could also have bearing on both the level of measured inequality and its evolution over time.

All our inequality measures, and other statistics, assume equality within the household (in terms of income per person, or income per equivalent single person), and are household size weighted. The Gini indices were calculated by numerical integration using the trapezoidal rule) of the empirical (household-level) Lorenz curve.

5 Results on the overall level of inequality

Figure 1 plots the proportionate change in income after all out data revisions against the log of original (unadjusted) income for 1985 and 1990. The fitted line was obtained by locally-weighted smoothing (using the “KSM” program in STATA). The figure also gives the fitted values for the increase in income attributable to grain revaluation alone, as well as the remainder due to other changes noted above. (The scatter of data points is for the total income increase due to data revisions.)

The proportionate change due to our data revisions tends to decrease as income increases, indicating that inequality falls after the changes. It is clear from Figure 1 that the revaluation of grain income in kind accounts for the bulk of the change, although the other changes are also inequality reducing on their own. The revaluation rates tend to be higher in 1990 than 1985, largely reflecting the rising divergence between market prices and planning prices.

Table 1 gives our estimates of two measures of income inequality over time. Both measures show rising inequality over the period for all three definitions of income. The magnitude of the increase is markedly less when one combines the new valuation methods with the new cost of living deflator. This can be seen more clearly from Figure 2. The adjustments to the data entail lower inequality, and a lower rate of increase in inequality.

The finding of lower inequality when our revisions are made to the income data is robust to the choice of inequality measure. This can be seen in Figure 3 and 4, which give the Lorenz curves before and after the data revisions, for 1985 and 1990 respectively. Figures 5 and 6 give the Lorenz curves for 1985 and 1990, based on the original and revised incomes (using the new valuation method, and new deflator). There is Lorenz dominance in both cases, so the conclusion that inequality increased is also robust to the inequality measure used. With the revisions to the primary income data, however, the two Lorenz curves have clearly converged considerably.

Figure 7 allow us to examine the effect on the inequality comparisons of introducing an allowance for scale economies. Instead of income per capita we use income divided by n^q where n is household size and q is a parameter between 0 and 1 interpretable as (minus one times) the elasticity of the cost of living with respect to household size. The conclusion that the Gini index of income inequality rose over the period is robust to the choice of q . So is the conclusion that the inequality is lower after making our revisions to the raw data, at any given value of q .

6 Inequality decompositions

Our aim here is not to attempt an exhaustive explanation of inequality in rural China, but rather to test sensitivity to the measurement problems. For this purpose, we decompose income into the 14 sources in Table 2. These are largely self-explanatory. The category “joint costs” allows for costs which we cannot apportion between factor income components. Table 3 gives the average shares of income attributed to these 14 sources. As expected, the revaluation of grain-income in kind entails a sizable increase in the share of income attributed to this component. On average, 21% of SSB’s income measure is attributed to grain, while this rises to 31% on revaluing

at average local selling prices. The new income component for imputed rents accounts for about 7% of income on average.

Tables 4 gives the source decomposition of the levels of inequality at the beginning and end of the period for the three income definitions. When compared to the original incomes, the new valuation methods entail a sizable increase in the share of inequality attributed to grain incomes, from 6% to 14% in 1990. Notice that, while the new valuation methods indicate lower inequality (Table 3), they also indicate that grain income is more covariate with total income, and hence it is found to account for a higher share of the (lower) level of inequality. These two findings are consistent. On the one hand, grain income from own production accounts for a larger share of the incomes of the poor, and this is why its revaluation tends to reduce measured inequality. On the other hand, better-off rural households tend to have higher incomes from grain production (even though the share of income from this source is lower). Since the undervaluation is in the output price, it acts like a constant proportionate mark-down of this component as in section 3, where it was shown that the undervaluation of grain income will then lead to an underestimation of the share of inequality attributed to this component of income as long as that share is less than 0.5, as is the case here (see the figures for grain in Table 4).

Table 4 also gives the shares of the measured increase in both inequality measures which are attributed to each component of income. Over the period, the revaluation of own-grain consumption entails a large increase in the share of rising inequality which is attributed this component. (It is readily verified from Tables 1 and 4 that for grain the second inequality in (12) holds for the Gini index but the first does not, although the difference is small; both inequalities in (12) hold for grain when using the log deviation.) The usual income definition used in data for

China suggests that income from collectives (including TVEs) was the most important single factor in the increase in overall inequality (Table 4). Our definition points instead to grain income.

Using our revised incomes (both revaluing grain income and using the new deflator) our results indicate that 104% of the increase in the Gini index over the period 1985-90 can be attributed to grain income; 61% was attributable to income from collectives (including TVEs). Smaller positive contributions to rising inequality came from self-employment in industry and construction (42%), labor earnings (36%) and services (32%). Against these positive contributions to rising inequality, there were large inequality-decreasing effects from private transfers (-131% of the increase in inequality) and other farm income (-65%).

Turning to the decomposition in terms of assets, we postulate that incomes are determined by the variables given in Table 5. The dependent variable is income per person, in constant prices. “Fixed productive assets” comprise the survey valuations of all immobile productive farm assets, expressed in constant prices using the same deflator as the dependent variable, and normalized by household size. “Labor force per capita” is the number of able-bodied workers per capita in the household. The variables “hilly area” and “mountainous area” are dummy variables for the geographic area in which the household lives, and the omitted dummy variable is that for households living on the plains. “Cultivated land,” “hilly land” and “fishpond” land are all areas of land owned or contracted per person in the household. The education variables are all dummy variables for the highest level of education reached by the workforce in the household; the omitted dummy variable is for a household in which all members are illiterate. We also include household size as an independent variable, to allow for possible scale economies.

Table 5 gives the regression coefficients for 1985 and 1990, for both SSB’s original

incomes and our adjusted incomes (revaluing grain and using the new deflator). The signs are unsurprising, and almost all coefficients are significant. By both measures, the income gain from higher fixed productive assets fell over the period 1985-90, while the returns to land and schooling (except college) rose.

Table 6 gives the simple correlation coefficients between each explanatory variable in the regressions and total income per capita. These will help in interpreting the inequality decompositions in Table 7 (analogous to Table 4).

A large share of the measured inequality at one date is attributable to the regression residual (Table 7); the values of R^2 in Table 6 are not unusually low for household-level cross-sectional regressions of this sort, but the unexplained component of the variance in incomes still accounts for 70-80% or more of the level of inequality. The residuals also account (positively) for a share of the change in inequality over time.

In terms of the asset decomposition, the biggest quantitative difference between the two income measures is in the estimated contribution of fixed farm assets to the change in inequality. Both measures suggest that this was inequality-reducing over the period. This is largely attributable to this factor's declining regression coefficient; the proportionate drops in the β coefficient on fixed productive assets in the income regressions (Table 5) are roughly the same as the drops in the shares of inequality (Table 7). Thus, the key factor appears to have been the lower "rate of return" to farm assets in 1990 than 1985. One might conjecture that wider access to capital in rural China during the 1980s helped reduce its returns. The reason why SSB's original incomes appear to have underestimated the (inequality-reducing) contribution of wider access to physical capital is that SSB's income measures underestimated the rate of return to these assets in the base period. This is

undoubtedly due to the undervaluation of grain income, leading to an underestimation of the marginal product of the farm capital stock.

Living in a mountainous area (relative to the plains, where farm land tends to be of better quality) was an important factor in explaining the level of inequality and an important source of rising inequality over time. Access to cultivated land was of negligible consequence for the level of inequality, but our adjustments to SSB's original incomes suggest that access to farm land was a more important source of higher inequality over time than one would have otherwise thought. This too is attributable in part to the increase in returns to land indicated by our corrections to the primary data. (Notice the large increase between 1985 and 1990 in the regression coefficient on cultivated land in Table 5, when based on our revised incomes; by contrast the original incomes indicate a small drop.) With our data revisions the correlation between land and income also increased (Table 6), adding further to its contribution to inequality. The revaluation of grain income in kind is clearly the main factor here too.

Both income measures (with and without our corrections) indicate that living in the mountains versus the plains was an important source of inequality, and a very important factor in the increase in inequality. Indeed, the distribution of households between mountainous areas and the plains accounts for 52% of the increase in the Gini index using our adjusted incomes (33% using SSB's original incomes.) Although fishponds only accounted for less than 2% of the level of inequality in 1990, they accounted for a sizable share of the increase in inequality, reflecting both a higher β in 1990 than 1985 (Table 5) and a higher correlation with total income (Table 6).

The importance of the geographic variables to how distribution evolves over time is consistent with the results of Jalan and Ravallion (1997) on these data. They found that rates of

consumption growth over time at the farm-household level are strongly influenced by geographic variables, controlling for household characteristics. This can be interpreted as a “geographic poverty trap” arising from the combined effect of credit market failure and an adverse effect of mountainous terrain and other geographic variables on the productivity of private investment.

Primary education was inequality-reducing, while other levels of education had the opposite effect, although the contribution was small in all cases (negligible in the case of college). Recall that we find increases in the β 's for the (non-college) education variables in Table 5. Lack of schooling beyond primary is negatively correlated with income (Table 6), so the higher returns put downward pressure on inequality. By contrast, the large increase in the returns to high school education put upward pressure on inequality, although this effect was dampened by an improvement in the distribution of high school education; the correlation coefficient fell slightly (Table 6) and the standard deviation also fell (by 7%).¹¹ In terms of the impact on inequality, a more equal distribution of secondary schooling helped compensate for its higher rate of return.

7 Conclusions

Tabulations of the distribution of rural incomes provided in the Statistical Yearbooks for China suggest a large increase in inequality after the mid-1980s. However, there are a number of

¹¹ Recall that the share of inequality attributed to any income determinant is the product of three things: the partial regression coefficient of income on that determinant, the simple correlation coefficient with income, and the ratio of the standard deviation of that determinant relative to the standard deviation of income.

concerns about the data underlying these numbers, as well as the level of their aggregation. While China's rural economy has been going through a structural transition, the processing methods used in the available survey data have not fully reflected those changes. Income in kind from the consumption of farm products has been systematically undervalued in official sources, due to a large and rising divergence in the 1980s between the prices used in official valuations and actual selling prices. Another concern is that existing data sources have ignored spatial differences in the cost of living, and how these have changed over time. Before we can be confident that there is in fact rising income inequality, these concerns should be addressed. Thankfully, one can go a long way toward fixing the main problems if one has access to the raw data from China's Rural Household Survey, which appear to be of good quality by international standards.

We find that about two thirds of the proportionate increase in measured income inequality in rural southern China between 1985 and 1990 vanishes once one revalues own-grain production at average local selling prices, imputes rents for housing and consumer durables, and allows for inter-provincial cost of living differences. After making these changes in the measured incomes at household level, instead of the 16% increase in the Gini index of income inequality between 1985 and 1990 implied by past data, we find a 6% increase; instead of a 36% increase in the average proportionate deviation from the mean, we find a 12% increase.

The undervaluation of income in kind from foodgrain production in the official data is the main source of bias in past inequality measures. This component of income tends to account for a higher share of the incomes of the poor, so its undervaluation leads to an overestimation of the level of inequality. Furthermore, the prices used by the provincial statistics offices diverged progressively over time from market prices, with the result that the undervaluation also leads to an

overestimation of the rate of increase in inequality.

What accounts for the measured increase in inequality not accountable to these data problems? The explanation depends on the income definition used. The revaluation of income in kind from grain production indicates that a much larger share of the (albeit smaller) increase in rural inequality was due to grain income than past data would have suggested. The income definition used in past work suggests that differing fortunes in grain production account for 15% of the rise in the Gini index; on revaluing grain income-in-kind at actual selling prices we find that this income component accounted for 58% of the increase in the Gini index. Private transfers were the strongest inequality-reducing factor amongst the income components we have measured.

We also estimated a decomposition of inequality in terms of land and (physical and human) assets. This suggests that the differences in income between those living in mountainous rural areas and those on the plains have been an important source of rising inequality, while diminishing returns entailed that the distribution of farm assets was inequality-reducing, once our corrections are made to measured incomes. Higher returns over time to good quality agricultural land (including whether one lives in the plains or the mountains) were inequality-increasing, even though the distribution of land was of little consequence to the level of inequality at any one date.

There are still problems in the data that we have not been able to deal with here, and at present it is only possible for us to perform these calculations for rural areas of four provinces. Nor have we addressed two other potential sources of rising inequality nationally, namely inequality between urban and rural areas, and inequality within urban areas. There are a number of as yet unresolved issues here, not least of which is allowing for differences in the cost of living between urban and rural areas (adjusting for inflation over time using separate urban and rural consumer

price indices does not incorporate the spatial difference at any one point in time). A further problem is obtaining a definition of income which is comparable between urban and rural areas of China; the rural and urban household surveys for China are largely independent and there appear to be a number of inconsistencies. Reprocessing of the raw survey data for both the urban and rural household surveys for all provinces could go a long way toward dealing with these issues.

References

- Atkinson, A. B. (1970). 'On the measurement of inequality' *.Journal of Economic Theory,* Vol. 2, pp. 244-263.
- Chen, S. and Ravallion, M.(1996). 'Data in transition: Assessing rural living standards in southern China'. *China Economic Review*, Vol. 7, pp. 23-56.
- Cowell, F. A. and Victoria-Feser, M. (1996). 'Robustness properties of inequality measures'. *Econometrica*, Vol. 64, pp. 77-101.
- Fields, G. S. (1996). 'Accounting for Differences in Income Inequality'. Mimeo, Cornell University.
- Howes, S. and Hussain, A. (1994). 'Regional Growth and Inequality in Rural China'. STICERD Discussion Paper 11, London: London School of Economics.
- Jalan, Jyotsna and Martin Ravallion (1997). 'Spatial Poverty Traps?', Policy Research Working Paper 1862, World Bank, Washington DC.
- Jenkins, S. P. (1995). 'Accounting for Inequality Trends: Decomposition Analysis for the UK, 1971-86'. *Economica*, Vol. 62, pp. 29-64.
- Khan, A. R., Griffin, K., Riskin, C. and Zhao, R. (1993). 'Sources of Income Inequality in Post-Reform China'. *China Economic Review*, Vol. 4, pp. 19-35.
- Knight, J. and Song, L. (1993). 'The Spatial Contribution to Income Inequality in Rural China'. *Cambridge Journal of Economics*, Vol. 17, pp.195-213.
- Lanjouw, P. and Ravallion, M. (1995). 'Poverty and household size'. *Economic Journal*, Vol. 105, pp. 1415-1434.
- Milanovic, B. (1996). 'Poverty and inequality in transition economies: What has actually

- happened' in Bart Kaminski (ed.) *Economic Transition in Russia and the New States of Eurasia*, New York: M.E. Sharpe.
- Pfingsten, A. (1988). 'Progressive Taxation and Redistributive Taxation: Different Labels for the Same Product?'. *Social Choice and Welfare*, Vol.5.
- Ravallion, M. and Chen, S. (1997). 'What can new survey data tell us about recent changes in distribution and poverty?'. *World Bank Economic Review*, Vol. 11. pp. 357-382.
- Rozelle, S. (1994). 'Rural Industrialization and Increasing Inequality: Emerging Patterns in China's Reforming Economy'. *Journal of Comparative Economics*, Vol. 19. pp. 362-391.
- Sah, R. and Stiglitz, J. E. (1992).. 'Peasants Versus City-Dwellers'. *Taxation and the Burden of Economic Development*. Oxford: Oxford University Press.
- Shorrocks, A. F. (1982). 'Inequality Decomposition by Factor Components'. *Econometrica*, Vol. 50. pp. 193-211.
- World Bank, (1992). '*China: Statistical System in Transition*'. World Bank, Washington DC.

Table 1: Income inequality measures for southern China

| | Inequality measure (%) | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 |
|---|---------------------------|-------|-------|-------|-------|-------|-------|
| 1. Original measure of household net income, as used in the Statistical Year-book for China | Gini | 29.11 | 30.12 | 31.04 | 32.96 | 33.75 | 33.90 |
| | Log deviation | 13.97 | 14.92 | 15.82 | 17.84 | 18.82 | 18.96 |
| 2. With new valuation methods (for income from own-grain production and imputed rents) | Gini | 27.47 | 28.75 | 29.46 | 30.57 | 31.10 | 30.88 |
| | Log deviation | 12.38 | 13.55 | 14.16 | 15.26 | 15.85 | 15.50 |
| 3. New valuation methods plus new cost-of-living deflators | Gini | 27.06 | 28.27 | 28.34 | 28.12 | 28.03 | 28.72 |
| | Log deviation | 12.02 | 13.11 | 13.14 | 12.96 | 12.93 | 13.43 |

Table 2: Components of rural household income

1. **Collective:** Income from collective businesses (collective united accounting units, TVE, economic union)
 2. **Grain:** Grain income
 3. **Non-grain farm:** Non-grain farming income
 4. **Animal:** Income from animal husbandry
 5. **Other farm:** Forestry, fishery, handicrafts and gathering & hunting
 6. **Industry:** Income from industry and construction (small business or self employer)
 7. **Services:** Income from transportation, commerce, restaurant, service and other (small business or self employer)
 8. **Labor:** Private sector labor earnings
 9. **State:** Wages and pensions from state or collective own enterprises
 10. **Public transfers:** Public transfers (government or village subsidies, bonuses and disaster release funds)
 11. **Private transfers:** Gifts or remittances from family members or relatives living outside the village (in rural or urban areas) for more than six months per year (those living outside the village for less than six months are counted as household members)
 12. **Other income:** Other factor income
 13. **Rent:** Imputed rent on housing and durable goods
 14. **Joint costs:** Joint costs of production which cannot be apportioned (tax, contract fee and the depreciation of fixed productive asset)
-

Table 3: Average income by source

| | Shares of 6-year mean income (%) | | |
|-------------------|----------------------------------|----------------------------|---|
| | Original incomes | With new valuation methods | New valuation methods plus new deflator |
| Collective | 3.01 | 2.37 | 2.26 |
| Grain | 21.12 | 31.11 | 31.48 |
| Non-grain farm | 22.27 | 17.58 | 17.61 |
| Animal | 17.82 | 14.03 | 14.09 |
| Other farm | 11.05 | 8.75 | 8.75 |
| Industry | 4.22 | 3.31 | 3.27 |
| Services | 7.31 | 5.75 | 5.68 |
| Labor | 8.83 | 6.91 | 6.78 |
| State | 2.43 | 1.92 | 1.93 |
| Public transfers | 2.74 | 2.16 | 2.15 |
| Private transfers | 2.97 | 2.36 | 2.30 |
| Other income | 1.40 | 1.10 | 1.09 |
| Rent | n.a. | 6.72 | 6.68 |
| Joint costs | -5.17 | -4.07 | -4.06 |
| | 100.00 | 100.00 | 100.00 |

Table 4: Factor decomposition of income inequality

| | Original incomes | | | | New valuation methods | | | | New valuations + | |
|-------------------|------------------|--------|---------|--------|-----------------------|--------|---------|--------|------------------|--------|
| | 1985 | 1990 | 1985-90 | | 1985 | 1990 | 1985-90 | | 1985 | 1990 |
| | | | Gini | LD | | | Gini | LD | | |
| Collective | 7.20 | 12.57 | 45.20 | 27.60 | 6.43 | 10.23 | 40.81 | 25.29 | 6.40 | 9.92 |
| Grain | 4.42 | 5.97 | 15.42 | 10.32 | 8.34 | 13.83 | 58.12 | 35.65 | 8.77 | 14.25 |
| Non-grain farm | 17.31 | 19.74 | 34.52 | 26.55 | 15.46 | 16.48 | 24.71 | 20.53 | 15.51 | 16.63 |
| Animal | 12.37 | 11.16 | 3.83 | 7.78 | 11.04 | 9.84 | 0.12 | 5.05 | 11.04 | 10.33 |
| Other farm | 14.52 | 9.77 | -19.08 | -3.52 | 11.97 | 7.69 | -26.82 | -9.31 | 11.70 | 7.25 |
| Industry | 4.84 | 8.41 | 30.10 | 18.40 | 4.27 | 6.75 | 26.71 | 16.58 | 4.39 | 6.58 |
| Services | 12.30 | 14.30 | 26.46 | 19.90 | 10.42 | 11.50 | 20.14 | 15.75 | 10.57 | 11.80 |
| Labor | 7.15 | 10.10 | 28.02 | 18.36 | 6.32 | 8.67 | 27.62 | 18.00 | 6.22 | 7.93 |
| State | 3.43 | 3.29 | 2.44 | 2.89 | 3.03 | 2.70 | 0.05 | 1.40 | 3.10 | 2.99 |
| Public transfers | 5.20 | 3.77 | -4.88 | -0.21 | 4.51 | 3.19 | -7.42 | -2.04 | 4.43 | 3.20 |
| Private transfers | 13.14 | 3.86 | -52.53 | -22.12 | 11.86 | 3.35 | -65.24 | -30.44 | 11.63 | 3.40 |
| Other income | 2.42 | 2.76 | 4.84 | 3.72 | 2.05 | 2.29 | 4.23 | 3.24 | 2.09 | 2.25 |
| Rent | n..a. | n..a. | n..a. | n..a. | 8.18 | 8.42 | 10.34 | 9.36 | 8.07 | 8.28 |
| Joint costs | -4.27 | -5.69 | -14.32 | -9.67 | -3.89 | -4.94 | -13.36 | -9.09 | -3.90 | -4.80 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

Note: The figures under "1985" and "1990" give the factor decomposition of the level of inequality. The figures under "1985-90" give the factor decomposition of the change in inequality using the Gini index and mean log deviation (LD).

Table 5: Regressions for real income per capita

| Variable | Original income | | With new valuation methods and new deflator | |
|--|---------------------|---------------------|---|---------------------|
| | 1985 | 1990 | 1985 | 1990 |
| Intercept | 333.02 (23.52) | 420.33 (23.99) | 401.91 (27.60) | 463.38 (29.82) |
| Fixed productive assets per capita | 0.32 (24.36) | 0.18 (17.01) | 0.36 (26.35) | 0.21 (19.47) |
| Household size | -14.96 (-12.33) | -21.30 (-14.38) | -16.48 (-13.21) | -24.08 (-18.33) |
| Household labor force per capita (able to work, if not working) | 155.69 (12.98) | 126.55 (9.91) | 181.50 (14.71) | 147.89 (13.05) |
| Hilly area (dummy variable for the locality of the household) | -74.48 (-12.22) | -101.67 (-13.42) | -93.30 (-14.88) | -89.485 (-13.32) |
| Mountainous area (dummy variable, as above) | -144.39 (-24.48) | -201.73 (-29.43) | -175.19 (-28.88) | -191.74 (-31.57) |
| Owned cultivated land area per capita | 0.14 (6.90) | 0.13 (3.77) | 0.17 (8.41) | 0.33 (10.89) |
| Area of hilly land per capita | -0.01 (-1.45) | -0.002 (-0.15) | -0.004 (-0.92) | 0.01 (0.80) |
| Area of fishpond land per capita | 0.08 (2.96) | 1.27 (14.40) | 0.07 (2.53) | 0.89 (11.40) |
| Highest education level is ...primary school | 38.00 (3.87) | 50.90 (3.93) | 36.64 (3.63) | 49.56 (4.32) |
| ...middle school | 78.19 (7.93) | 106.90 (8.30) | 76.48 (7.55) | 97.38 (8.53) |
| ...high school | 117.57 (10.75) | 171.63 (12.13) | 115.55 (10.27) | 148.20 (11.81) |
| ...technical school | 133.40 (4.89) | 216.59 (7.55) | 121.10 (4.32) | 193.13 (7.59) |
| ...college | 226.61 (3.43) | 253.68 (4.43) | 213.28 (3.14) | 203.78 (4.02) |
| R^2 | 0.185 | 0.210 | 0.217 | 0.247 |

Note: Monetary values for 1990 are in 1985 prices

Table 6: Correlation coefficients with total income per capita

| | Original income | | New valuation methods plus new deflator | |
|-------------------|-----------------|-------|---|-------|
| | 1985 | 1990 | 1985 | 1990 |
| Productive assets | 0.24 | 0.18 | 0.27 | 0.23 |
| Household size | -0.04 | -0.05 | -0.05 | -0.05 |
| Labor | 0.20 | 0.15 | 0.21 | 0.19 |
| Hilly area | 0.05 | 0.08 | 0.04 | 0.09 |
| Mountain | -0.24 | -0.29 | -0.27 | -0.30 |
| Cultivated land | 0.08 | 0.05 | 0.09 | 0.13 |
| Hilly land | 0.00 | -0.05 | 0.00 | -0.03 |
| Fishpond | 0.06 | 0.16 | 0.05 | 0.13 |
| Primary school | -0.12 | -0.13 | -0.12 | -0.12 |
| Middle school | 0.06 | 0.05 | 0.06 | 0.05 |
| High school | 0.13 | 0.13 | 0.13 | 0.10 |
| Technical school | 0.03 | 0.05 | 0.03 | 0.05 |
| College | 0.01 | 0.03 | 0.01 | 0.03 |
