1994

To Have and Have Not: Explaining Inequality in Rural China

Jonathan Morduch

Terry Sicular

POLITICAL
ECONOMY
RESEARCH
GROUP

PAPERS IN POLITICAL ECONOMY

Paper No. 51

"To Have and Have Not: Explaining Inequality in Rural China"

Jonathan Morduch and Terry Sicular

The UNIVERSITY *of* WESTERN ONTARIO

The Political Economy Research Group was established in the faculty of Social Science at the University of Western Ontario in 1988. Its purpose is to foster scholarship, teaching and interdisciplinary research in political economy, with a focus on:

1. the application of economic models and methods to the study of political processes and institutions,
2. the economic impact of political processes and institutions,
3. the influence of economic factors on the formation of public policy and on institutional change,
4. the politics of economic policy making,
5. the political, social, and economic effects of public policy.

For further information:
Political Economy Research Group,
Department of Economics,
Social Science Centre,
London, Ontario, Canada N6A 5C2
phone: (519) 661-3877
fax: (519) 661-3292

ISSN: 1181-7151
ISBN: 0-7714-1716-0

TO HAVE AND HAVE NOT: EXPLAINING INEQUALITY IN RURAL CHINA*

Jonathan Morduch
Harvard University

Terry Sicular
University of Western Ontario

November 20, 1994

*This paper contains work in progress, and the findings should be viewed as preliminary. We have benefited from discussions with Joseph Stern, Sudhir Anand, and Robin Cowan. We are grateful for the research assistance of Sarah Cook and Joseph Zveglich. Support for this project was provided by the National Science Foundation.

I. Introduction

Economic development brings changes in both the average level and the distribution of income. The nature of these changes is the subject of some debate. Recent studies have challenged the Kuznets (1955) inverted-U hypothesis that inequality first rises and then falls with income growth.[1] Questions have also been raised about the causal relationship between growth and inequality. Through what mechanisms does growth influence inequality? Which factors explain income dispersion? Are certain groups clustered in particular parts of the income distribution? Does inequality itself affect the pace and pattern of growth?[2]

Economic reform or transition in socialist economies also raises questions about distribution. Here the main issue is not whether inequality rises during the process of reform, as most observers agree that some rise in inequality is inevitable. Rather, the key question is who benefits more and who benefits less, that is, what explains the pattern of inequality that emerges. One view is that the elite under socialism have special power and connections, and that these advantages enable them to benefit disproportionately from the reforms. If so, then inequality should be linked to political position and class status. Another view is that by opening up markets and providing new opportunities, the reforms benefit those who were less privileged under socialism. If this second view is correct, then inequality should be largely explained by the usual sorts of individual and household characteristics such as education, dependency ratios, age, and so on.[3]

Properly addressing such questions involves disentangling the sources of income inequality. The method commonly used to accomplish this task is inequality decomposition. Inequality decomposition takes an aggregate index of inequality and mathematically decomposes it into its

---

[1]See, for example, the recent studies by Anand and Kanbur (1993) and Fields and Jakubsen (1993).

[2]The literature on growth and income distribution, which was particularly active during the 1970's, has been re-activated lately in different forms. For example, Alesina and Rodrik (1994) provide evidence that equality aids growth through the political process.

[3]This debate is discussed in more detail below.

component parts. The decomposition is often carried out over population groups distinguished on the basis of regional, ethnic, educational or other characteristics.[4] Such an approach measures the contributions to overall inequality of income dispersion *within* groups versus that *between* groups, and inference follows accordingly.[5]

In this paper we reevaluate the usual approach to inequality decomposition. Certain drawbacks of the standard method lead us to propose an alternative, "bottom-up" approach that in a sense turns the traditional methodology on its head. The bottom-up strategy involves using econometric estimation to form conditional expectations of income and constructing inequality indices based on those expectations. This strategy has several advantages. Incorporating econometric estimation into the process of inequality decomposition allows consideration of multiple, continuous explanatory variables; endogenous relationships; and tests of statistical significance in the explanation of income patterns.[6] Moreover, it allows flexibility in handling small samples. The standard, top-down approach can be seen as a special case of the bottom-up approach, and this makes comparisons to existing studies straightforward.

The first step in the bottom-up approach is consideration of the systematic versus non-systematic elements which explain variation in income using econometric analysis (such issues are the essence of much econometric work). Consideration of systematic components leads naturally to a host

---

[4]Alternatively, decomposition can be done on the basis of income source (e.g., wage income, transfer income, capital income).

[5]Standard practice in inequality decomposition is first to choose an aggregate inequality index which satisfies the requisite properties (e.g., scale independence and the Pigou-Dalton Transfer Principle). Attention must be limited to the set of indices which are decomposable, such as Theil's two entropy-based measures and the variance of the logarithm of income (to be strictly decomposable, the index needs to be additively separable). The sample is then divided into disjoint groups (say, black and white, Chinese and Malay, literate and illiterate). *Between*-group inequality is the level of aggregate inequality which would exist if all dispersion of income within groups was suppressed. *Within*-group inequality is the level of aggregate inequality which would exist if group means were equalized and all variation came only from dispersion around group means. Comprehensive discussions of this approach are provided by Fields (1980) and Anand (1983).

[6]As described by Fields (1980), the use of ANOVA coupled with the VarLog inequality index provides some of these advantages as well.

of issues surrounding specification and identification. Consideration of non-systematic components leads naturally to investigation of forms of heteroskedasticity. Once the variables that determine income have been identified and their relationship with income has been estimated, we proceed with decomposition. As usual, the decomposition requires choice of an inequality index that attaches social weights to different aspects of the income distribution. In decomposing this index, definition of groups is now informed by the econometric results.

We apply the bottom-up approach to a new set of data for a sample of farm households in Zouping, a county in central Shandong Province that is fairly representative of rural areas in northern China.[7] Zouping County has seen annual growth of per capita income at rates of 8 to 10 percent during the 1980s and early 1990's, and the data give a detailed snapshot of inequality in a typical rural area in rapid transition. The effect of the reforms on inequality in rural China has been widely discussed in the literature, with frequent use of inequality decomposition [Hare (1994), Hussain *et al* (1991), Khan *et al* (1992), Knight and Song (1993a), and Rozelle (1993)]. Our analysis complements these existing studies, some of which look at particular localities [Putterman (1993) and Hare (1994)], and some of which use nationwide surveys or aggregate information to examine inequality over dispersed regions [Griffin and Zhao (1993)].

Our findings show that both status variables and the usual sorts of household characteristics contribute to inequality. Status variables together contribute about one-third of overall inequality in our sample. The usual sorts of household characteristics together explain about 40 percent of overall inequality. Thus if Zouping County is representative, both sides of the debate on what explains inequality in transitional economies have some relevance for rural China.

Our findings also demonstrate that qualitatively different conclusions can emerge when using traditional accounting-based (top-down) approaches versus the regression-based (bottom-up) approach. For example, standard decomposition overstates the contribution of place of residence.

---

[7]Appendix table 1 gives descriptive statistics for Zouping County, with comparisons to national averages. For a more detailed description of the county and survey sample see Sicular (1993).

More generally, standard decomposition masks the multiple and potentially complex relationships that underlie inequality, whereas working from the bottom up allows a fairly rich picture to emerge. Thus we find that households with Party members can be divided into two groups. One group has done well, taking advantage of special opportunities to set up and profit from household sideline businesses. The other group has not--indeed, it has done no better, and in some cases worse, than average. The contribution of Party membership to inequality given by bottom-up decomposition reflects these distinctions, but also depends on the distribution of Party members in the sample. Similarly, we find that even though land per capita is distributed fairly evenly through the sample, it explains a substantial share of inequality. This occurs because land has a large impact on the level of income, and so a small amount of inequality in landholdings translates into a large amount of inequality in income.

II. Decomposition Methodology

The Top-Down Approach

We begin by considering a simple decomposition using the Zouping County data. For the sake of exposition, we will discuss decomposition using the Theil-L index of inequality, although we also present results for the Theil and the variance of the logarithm of income. Choice of inequality index will, of course, depend on the aspects of inequality that one wants to emphasize. The Theil measures are more sensitive to dispersion at the upper tail of the income distribution, while the variance of log income exhibits sensitivity to differences in incomes of poorer individuals.

The Theil-L index is relatively simple and relates income shares of groups to their population shares. The formula for the Theil-L is

$$TL = \sum_{i=1}^{N} \frac{n_i}{N} \log\left(\frac{Y/N}{y_i/n_i}\right) \qquad (1)$$

where Y is the total income for the whole population, N is number of people in the population, and y_i and n_i are the analogues for individual i (for an individual, $n_i = 1$). This formula can be rewritten

more simply as the difference between the logarithm of average income and the logarithm of the geometric mean of income μ:

$$TL = \log(Y/N) - \log(\mu). \tag{2}$$

In the standard decomposition of inequality over groups, J groups are defined, and the above formula is decomposed into within- and between-group components. The between-group component resembles the overall formula for the Theil-L, but group incomes and populations are substituted for the individual values:[8]

$$TL_B = \sum_{j=1}^{J} \frac{N_j}{N} \log\left(\frac{Y/N}{Y_j/N_j}\right). \tag{3}$$

This expression gives the amount of total inequality which is due to differences between the average incomes of groups, while suppressing all intra-group income variation. Within-group inequality is the amount of total inequality which arises if differences in average income between groups are suppressed, leaving just intra-group variation:

$$TL_W = \sum_{j=1}^{J} \frac{N_j}{N} \sum_{i \in j} \frac{1}{N_j} \log\left(\frac{Y_j/N_j}{y_i}\right). \tag{4}$$

Note that in this standard formulation inequality between groups depends on differences in the average incomes among groups. The average income of a group is the conditional expectation of income based on affiliation with that group. In addition, groups in top-down decomposition are out of necessity defined in limited ways. These features of top-down decomposition restrict the insights it provides about what truly drives inequality.

In applying the standard decomposition method to our Chinese data, we follow the recent literature on inequality in China by focusing on the roles of region and Communist Party membership during the reform period. Some observers have argued that by providing privileged access to

---

[8]For a derivation, see Appendix C of Anand (1983), pp. 329-30.

productive resources, information, and employment opportunities, Party membership or cadre status confers advantages which should show up in income levels. Others have emphasized the role of geographic location in explaining inequality. (These issues are discussed in more detail below.) We therefore choose to decompose inequality first between individuals in households with and without Party members, and then between individuals living in different locations.

TABLE 1
Inequality Indices: Role of Communist Party Membership

	Aggregate	Households with Party Members	Households with no Party Members
Average Annual Per Capita Income	1116 yuan	1353 yuan	1036 yuan
Income Share	1.00	0.31	0.70
Population Share	1.00	0.25	0.75
Theil-L	0.169	0.193	0.151
Theil-T	0.171	0.182	0.155
Variance of Log Income	0.337	0.408	0.301

NOTE: Calculations are for the sample of 1028 individuals in 254 households, weighted to correct for sampling bias across villages.

Table 1 gives information on levels of income and inequality for the sample as a whole and for the two groups (Party, non-Party). The aggregate Theil-L calculated for per capita household income over the sample of 1028 individuals in 254 households is 0.17. Individuals in households with Party members on average have substantially higher per capita income–1353 yuan versus 1036 yuan. Those in households with Party members receive 31 percent of total income in the sample while representing just 25 percent of the population.

All the same, decomposition shows somewhat surprisingly that most inequality is due to differences *within* the groups rather than between them. As shown in table 2, inequality between groups explains at most 4 percent, and inequality within groups 96 percent, of aggregate inequality.

In these calculations Party membership is the only variable used to distinguish among

individuals. Party membership, however, is just one determinant of income. Omitted variables such as age, education, and location could also cause some or all of the inequality between and within groups. In effect, then, this approach is analogous to running a regression which has a constant and a Communist Party membership dummy variable as the sole explanatory variables. The results of the simple decomposition are therefore in some sense biased and must be interpreted carefully: the independent role of Communist Party membership is impossible to discern. With a fully specified regression (or a more disaggregated decomposition), Party membership may emerge more or less prominently.

TABLE 2
Inequality Decomposition: Role of Communist Party Membership

	Aggregate Inequality	Between-Group Inequality	Within-Group Inequality
Theil-L	0.169	0.007 (4%)	0.162 (96%)
Theil-T	0.171	0.007 (4%)	0.163 (96%)
Variance of Log Income	0.337	0.010 (3%)	0.328 (97%)

NOTE: Calculations are for the entire sample, composed of 1028 individuals in 254 households, weighted to correct for sampling bias across villages. Numbers in parentheses are the percentages of aggregate inequality contributed by the between- versus within-group components.

To some extent the problem of omitted variables can be overcome by decomposing inequality over multiple groups, for example, over both location and Party membership. Table 3 presents the results of this exercise. Decomposition by village of residence alone shows that between-village inequality is substantial, accounting for about 40 percent of aggregate inequality. When decomposition is carried out over both Party membership and village, the share of between-group inequality rises by about 5 to 6 percent points. Between-group inequality in the two-way decomposition exceeds the sum of between-group inequality in the one-way decompositions, which implies that omitting variables indeed affect the results.

TABLE 3
Inequality Decomposition: Role of Location and Communist Party Membership

		Aggregate Inequality	Between-Group Inequality	Within-Group Inequality
By Village	Theil-L	0.169	0.071 (42%)	0.098 (58%)
	Theil-T	0.171	0.072 (42%)	0.099 (58%)
	Variance of Log Income	0.337	0.130 (39%)	0.207 (61%)
By Village and Party Membership	Theil-L	0.169	0.079 (47%)	0.090 (53%)
	Theil-T	0.171	0.081 (48%)	0.089 (52%)
	Variance of Log Income	0.337	0.147 (44%)	0.190 (56%)

NOTE: Calculations are for the sample of 1028 individuals in 254 households in 16 villages, weighted to correct for sampling bias across villages. Numbers in parentheses are the percentages of aggregate inequality contributed by between- and within-group components.

Discussion

The present example of the top-down approach raises several issues. First, sample size poses a constraint. Our sample contains only 16 households from each village, and in some villages only one or two households contain Party members. Consequently, several "groups" contain few individuals. The presence of small groups adds a great deal of noise to the calculations and so makes the decomposition problematic. As we discuss below, using regression analysis to form conditional expectations can help alleviate this problem.

Second, in order to decompose over just two factors (Party membership and village), we have broken the sample into 32 separate groups. This number of groups is unwieldy and exacerbates the small sample problems noted above. If we wish to take the logical next step and control for additional variables, the number of groups would quickly balloon. Just adding education could require dividing the sample into 96 groups (2 Party member categories *times* 3 educational categories *times* 16 villages). Handling multiple explanatory variables is clearly cumbersome.

Third, variables like education and, to a greater extent, age are not categorical but continuous.

Top-down decomposition can only be carried out over discrete groups.[9] In some cases continuous variables can be converted into discrete ones with little violence to the data, but this is not clear *a priori*.

Fourth, education and certain other variables may be a function of household income. In China the endogeneity of education is not yet an issue, because adult education levels were largely determined before the reforms and are probably independent of current incomes. Endogeneity may become an issue in the near future, however, once children educated after the reforms enter the labor force. The standard approach to decomposition ignores endogeneity: conditional expectations are not identified in a causal sense.[10]

Fifth, in using the standard approach to decomposition we have no simple measures of statistical significance, although (seldom employed) t-tests can be used to test the significance of differences in group means. Finally, the groups of interest in top-down decomposition are determined by the researcher *a priori*. In contrast, a regression-based approach allows the data to determine which factors are most important.

Despite these drawbacks, the traditional approach has certain appealing features. By not putting structure on the income relationship, this approach can accommodate heteroskedasticity in general forms. Furthermore, for certain policy questions we may not need to worry about some of the considerations mentioned above--knowing the rough relationships among groups may be sufficient. As discussed in the conclusion, however, the bottom-up approach often yields additional information that can be of considerable interest for policy purposes.

The Bottom-Up Approach

Like the standard approach, bottom-up decomposition is built around deviations from

---

[9]Using analysis of covariance methods (ANOCOVA) with the VarLog does allow consideration of continuous variables, although this approach is not very flexible.

[10]While they have other appeals, standard ANOVA and ANOCOVA methods fall short in handling endogeneity. The regression-based approach proposed here handles two-stage least squares (or other specifications) as a straightforward generalization.

conditional expectations of income. These conditional expectations are estimated econometrically. We therefore begin by specifying an income equation, with independent variables and a functional form that approximate what we believe to be the true relationship between income and its determinants. Suppose the relationship between per capita income and its determinants $x_1,...,x_k$ takes the form

$$y_i = A x_{1i}^{\beta_1} x_{2i}^{\beta_2} ... x_{ki}^{\beta_k} u_i \,, \tag{5}$$

where u_i is a stochastic term. Then the regression equation is log-linear:

$$\log(y_i) = \alpha + \beta_1 \log(x_{1i}) + \beta_2 \log(x_{2i}) + ... + \beta_K \log(x_{Ki}) + \epsilon_i \,, \tag{6}$$

where $A = e^{\alpha}$ and $u_i = e^{\epsilon_i}$. Estimation of this equation in and of itself provides a good deal of interesting information which cannot be obtained by decomposition alone. In particular, it quantifies the relative importance of the various factors that explain income levels and measures their statistical significance.

From here, decomposition can take one of two paths. The first path retains the standard focus on discrete groups such as households with and without Party members, where the choice of groups is driven by policy or other *a priori* considerations. Inequality between and within these groups is then calculated in the standard fashion described above. Both between- and within-group inequality so measured, however, contain both systematic and non-systematic components. The systematic and non-systematic components can now be calculated using the regression results. Thus one can determine how much of the inequality between Party and non-Party groups is due to differences between these groups in levels of education, village of residence, and so on, and how much is due to Party membership once other variables are held constant. Similarly, one can determine how much of within-group inequality reflects variation in levels of education, etc., among individuals in the groups, and how much is unexplained or non-systematic.

The second path, which is the path we take in this paper, lets the structure of the decomposition follow naturally from the econometric analysis. In so doing, we effectively redefine the

meanings of "between" and "within." Between inequality now refers to all systematic variation in income due to the unequal distribution of attributes like education or location. Within inequality now refers to non-systematic or residual variation in income. These definitions are natural extensions of the standard concept of groups: groups are now defined as sets of individuals who are identical not just in one attribute, but in all attributes (i.e., individuals with the same education *and* the same village of residence *and* the same ethnicity, etc.).

The decomposition is carried out directly on the individual attributes which determine income as shown in the regression equation above. Between-group inequality is still the amount of inequality due to differences in conditional expectations of income, but now the conditional expectations are values predicted by the regression rather than simple averages of income over broadly defined groups.

The exact formulae for calculating between- and within-group inequality will depend both on the functional form of relationship between income and its determinants and on the choice of inequality measure. If the functional form is non-linear, as it is here, then the formulae will necessarily include terms that reflect multiplicative interactions or covariation among different variables. These terms arise because inequality is decomposed additively, but the income equation is non-additive.

In the case of the Theil-L index, if the relationship between income and its determinants is that shown in (5), then inequality can be broken down into three components

$$TL = TL_B + TL_W + TL_\theta \quad . \tag{7}$$

The "between" component of inequality is now

$$TL_B = \sum_{j=1}^{K} \hat{\beta}_j \left(\log(\bar{x}_j) - \log(\mu_j) \right) \quad , \tag{8}$$

where hats indicate regression estimates, $\bar{x}_j$ are sample means of the independent variables, and μ_j are their geometric means. This expression shows that between inequality is the weighted sum of the

Theil-L indices of the K attributes, with weights given by the estimated regression coefficients.[11] The contribution of any single attribute to overall inequality thus depends in part on the size of its coefficient: dispersion in attributes with large coefficients counts more than dispersion in attributes with small coefficients. The contribution of an attribute depends also on how equally or unequally it is distributed, as measured by the inequality index. If all attributes are distributed evenly, that is, if members of the population are identical in all attributes, then between inequality will equal zero.

"Within" inequality is

$$TL_W = \log(\bar{\hat{u}}) - \log(\mu_{\hat{u}}) \quad . \tag{9}$$

Within inequality equals the value of the Theil-L index for $\hat{u}$, the stochastic element in income as calculated using the estimated residuals from the regression. If the regression fits the data perfectly, that is, if all points lie on the estimated regression line, then within inequality equals zero.

Since the income equation is non-linear, the decomposition includes a third term to adjust for multiplicative interactions among the independent variables:

$$TL_0 = \log\left(\frac{Y/N}{\hat{A}\bar{x}_1^{\hat{\beta}_1}\bar{x}_2^{\hat{\beta}_2}\ldots\bar{x}_k^{\hat{\beta}_k}\bar{\hat{u}}}\right) \quad . \tag{10}$$

The term in parentheses equals the ratio of average income to the income of the "average" individual.[12] This ratio equals one if all attributes are identical for all individuals, in which case TL_0

[11]Between inequality contains the contributions of all independent variables, including those with estimated coefficients that are not statistically significant. Using the regression results, one can calculate the standard errors of each variable's contribution to inequality and determine whether its contribution is statistically significant. Calculation of standard errors and discussion of significance for the Theil-L contributions will be included in future versions of this paper.

[12]Bottom-up decomposition using the Theil-T contains a similar term. Bottom-up decomposition using the variance of log income contains covariance terms. Multiplicative and covariance terms do not appear in standard decomposition because the population is broken into discrete groups which by definition are additive and have zero covariance.

equals zero.[13]

Discussion

One of the most appealing features of standard decompositions is that it imposes no form on the distribution of error terms. The bottom-up approach described above maintains this feature, with the additional advantage that so long as the regression specification is correct, heteroskedasticity can be ascribed just to the patterns of error terms and not to omitted variables as in the standard decomposition. To the extent possible, all systematic contributors to within-group dispersion have been isolated and quantified, leaving a residual term which solely reflects non-systematic factors. This advantage is predicated on the assumption that the income equation is specified correctly. If not, measured within-group inequality will partly reflect specification error. (This is a standard problem in decompositions, whether they are based on regression analysis or not.)

One important aspect of specification is correcting for endogeneity. Endogenous relationships can be handled naturally at the regression stage of bottom-up decomposition using two-stage least squares. The estimated coefficients from a two-stage least squares regression would simply be used in the formulae above.

Bottom-up decomposition allows us to decompose inequality over both continuous and categorical variables. The bottom-up approach is in fact the continuous analogue of standard, top-down decomposition. Standard decomposition can be viewed as a special case of bottom-up decomposition: standard decomposition over one attribute such as Party membership is equivalent to bottom-up decomposition using a linear regression equation that contains a dummy variable for Party membership as the sole independent variable.

---

[13]Note that if a variable enters the income equation in non-logarithmic form, for example, a dummy variable or the stochastic term, then the denominator is calculated using the average of e raised to its power.

III. Application of the Bottom-Up Approach to Rural China

Specification of the Econometric Model

Studies of income of rural households in developing countries usually identify household characteristics such as education, the number of adult workers relative to dependents, the number of male workers relative to the total family labor force, and the age of family members or stage in the family life cycle as independent variables which explain income. Family size can be important, especially if markets for labor or credit are incomplete. In the absence of well-developed factor markets, larger or extended households may be better able to garner sufficient resources to carry out certain types of production or investment. Endowments of key physical assets such as land, and the attributes of those assets (such as land quality, fragmentation, etc.), can also influence income.[14]

These household characteristics appear as independent variables in our regressions, which contain the log of average years of education per adult (logeduc), of average age of adults (logage), of family size (logpop), of cultivated land area per capita (loglandpc), and of the number of plots of land (logplots). The number of adults per capita (adultspc), proportion of adults that are male (maleadults), and the share of land that is hilly or mountainous (hillarea) also appear on the right hand side,[15] as well as a dummy variable (ddisaster) that indicates whether the household experienced natural disaster such as drought or hail in the year of the survey.[16]

One variable that is sometimes omitted in econometric studies of household income is the number of hours or days worked. In our sample days worked per capita varied substantially, ranging

---

[14]Simultaneity is a potential problem for some of these variables, but less so in China than in other countries. For example, households with higher incomes per capita might be expected to own more land per person. In China markets for farmland do not yet exist, and land is distributed by villages to households on a per capita or per worker basis. Thus land area per capita is largely determined by village of residence (and will thus be reflected largely in the village fixed effects).

[15]Adults are defined as all family members aged 18 through 65. Note that the regressions were run using alternative measures of education such as education of household head and maximum years of education among adult family members. The results of these alternative regressions did not differ substantially from those reported here.

[16]See the appendix for a list of variable names and definitions.

from 61 to 415 days (days are standardized to equal eight hours of work time). Such variation can cause differences in per capita income. Estimates of earning functions for individuals in the labor literature usually hold time worked constant either by using earnings per hour as the dependent variable or by including time worked as an independent variable. We adopt the latter approach and include days worked per adult (workdays) on the right-hand side.[17]

Geographical location is often discussed in the context of rural inequality, especially in China where geographic mobility has been limited and geographical diversity is substantial. Within a county such as Zouping, village of residence is probably the key location variable. Land quality, water conditions, and distance to markets vary considerably among villages. Villages also differ in their leadership and paths of development. Some villages, for example, have energetic leaders who have successfully promoted collective village enterprises. These enterprises provide employment for village members and supply funds that can be used to support local public services such as education and infrastructure investment. Other villages lack effective leadership. When villages without effective leadership prosper, they have usually done so by following a development path based on private household enterprise. Since village characteristics play a role in determining income, the estimated regressions contain village dummy variables.

A major topic of debate for China and other socialist (or formerly socialist) countries is whether inequality in the reform period arises primarily because of the usual household characteristics discussed above, or whether it arises because some households have special access or privileged status.[18] According to one side of this debate, the socialist elite (cadres and members of the Communist Party) can use its political connections and knowledge to advantage during the reform period [McAuley (1990), Oi (1989), and Shirk (1989)]. These connections and knowledge are valuable assets in an environment where information about and access to economic opportunities are hard to

---

[17]Days worked by family members could be endogenous, as in theory the amount of leisure consumed is a function of income. We have not corrected for endogeneity here in part because days worked include non-family labor, and in part because it is difficult to identify a suitable instrumental variable.

[18]See Róna-Tas and Kolosi (1993) for an excellent summary of this debate.

obtain.

The other side argues that households that were less privileged prior to the reforms benefit more from the reforms. As markets develop the power of the socialist elite erodes, because markets provide new opportunities for mobility and reward effort, education, and skill [Nee (1989) and (1991)]. Furthermore, the pre-socialist elite (former landlords and entrepreneurs) have experience and attitudes necessary to profit from the new economic environment [Szelényi (1988)]. Indeed, the discrimination endured under socialism by these and other "bad elements" may have sharpened their entrepreneurial skills.

If the former view is correct, then in China certain status-related variables should be included in the analysis. One important status variable is class background. Under the Communist regime families were given class labels depending on their economic situation before the revolution. Rural families were classified as landless peasant, poor peasant (dpoorpeasant), middle peasant, rich peasant (drichpeasant), and landlord (dlandlord).[19] While some debate exists regarding the degree to which these labels accurately correspond to the pre-1949 economic status of families, it is probably safe to say that there was some correlation between a family's class designation and its prior economic situation. Moreover, one's treatment after the revolution depended on one's class label. Individuals from landlord families were often denied access to educational opportunities and career advancement, while individuals from landless or poor peasant families were given special opportunities.

Membership in the Communist Party (dcommunist) and cadre standing are also relevant status variables. The term "cadre" refers to individuals who hold positions of political leadership in government or in collective or state enterprises (for example, the head of the village government or the director of a collective enterprise). Cadres are often but not always members of the Communist Party. By definition cadres hold paid positions that involve full- or part-time work.

---

[19]Terms in parentheses are the names of dummy variables indicating whether a household has this class background. Landless and poor peasant households are grouped together. To avoid multicollinearity, the dummy for middle peasant households is dropped. Here and below, variable names that begin with "d" are dummy variables.

Cadre households fall into three different categories. First, some households have had cadre members over the long term, both before and since the reforms (dcadre). Second, some households have only recently had members assume the role of cadre (dnewcadre). Third, some households contained cadres prior to the reforms, but household members no longer occupy such positions (doldcadre). Since the economic advantages of these three types of cadre households could differ, they are treated as distinct groups in the regression.

Some authors have suggested that in rural China access to off-farm employment may be limited and distributed unequally [Knight and Song (1993b); Hare (1994)]. Wage employment is usually in enterprises run collectively by village or township governments, and these jobs are often allocated by local Party members and cadres. Similarly, households wishing to establish or expand a family sideline enterprise require local permits and approvals. Such permits may be awarded on the basis of political or social criteria rather than the economic viability of the business. Under these circumstances households with members who hold wage jobs and families with sideline businesses may enjoy extra-normal rents. We therefore include dummy variables for households with such forms of employment (dwagejob and dsideline, respectively).

Finally, some determinants of income may work in combination. Here we focus on the possible interactions between Party membership and certain other variables. It may be, for example, that having a wage job or a sideline business does not, in and of itself, enhance household per capita income, but that households that contain a Party member may have access through their political connections to particularly remunerative jobs, or they may be better able to profit in private business. We include interaction terms for households that currently contain both Party members and cadres (dcomm&cadre), Party members and wage employment (dcomm&wagejob), and Party members and a family sideline business (dcomm&sideline).

In theory, then, income is a function of a range of variables, including the usual family characteristics such as age and education, location of residence, time worked, and indicators of special status such as family background, Party membership, and cadre status. The underlying functional

form of this relationship is not known, but it is likely to be non-linear. We present results for estimates of both the log-linear and semi-log functional forms, both of which are commonly used in the literature. For completeness we have estimated regressions using both per capita household income and total household income as the dependent variable (results for total household income appear in the appendix). Results for the different regression specifications are quite similar.

Estimation results

The regression results indicate that both the usual sorts of household characteristics and status variables influence income levels (tables 4 and 5). Coefficients for land, the number of adults per capita, and days worked per adult are positive and consistently significant. Thus individuals in households with more land or in households with lower dependency ratios tend to have higher per capita income, as do individuals in households with more days worked per adult. Land-related characteristics (plots, hillarea, ddisaster) are significant only sporadically and then usually only in regressions without village dummies. This probably reflects correlations between such characteristics and village of residence.

Education's coefficient is positive, but significant only in regressions without village dummies. This again suggests correlation with village of residence, which is perhaps not surprising because wealthier villages and townships are likely to devote more resources to local schools. Coefficients on family size, age, and share of adults that are male are in general insignificant.

Variables for family background are uniformly insignificant, which suggests either that family background is unimportant, or that the advantages and disadvantages of having "good" versus "bad" class background cancel out. The cadre variables are also insignificant, except for long-term cadres (dcadre) in regressions that exclude the wage job and sideline dummies (regressions 3 and 4). This result can be explained by the fact that cadres by definition hold wage jobs. In regressions that include a dummy for wage employment, its coefficient is significant and positive, while the cadre dummies are no longer significant. Thus if cadre families enjoy higher incomes, it is only because they have wage employment. Put differently, the per capita incomes of families with cadres are no

Table 4
Estimates of Per Capita Household Income,
Log-Linear Specification
(n=254)

	1	2	3	4	5	6
loglandpc	.669* (.121)	.498* (.179)	.566* (.126)	.456** (.186)	.662* (.123)	.497* (.179)
logplots	-.046 (.125)	.167 (.135)	.135 (.129)	.255*** (.141)	-.060 (.124)	.096 (.135)
hillarea	-.780 (.579)	-.106 (.644)	-1.125*** (.603)	-.148 (.663)	-1.018*** (.582)	-.712 (.682)
logpop	-.045 (.150)	-.166 (.142)	-.129 (.155)	-.180 (.145)	-.043 (.149)	-.113 (.142)
adultspc	.532* (.205)	.547* (.182)	.331 (.208)	.416** (.182)	.480** (.203)	.489* (.182)
maleadult	-.341 (.300)	-.141 (.263)	-.291 (.319)	-.025 (.276)	-.239 (.300)	-.008 (.265)
logage	.122 (.164)	-.169 (.147)	.095 (.172)	-.257*** (.153)	.145 (.162)	-.167 (.146)
logeduc	.235** (.107)	.052 (.099)	.282** (.113)	.037 (.104)	.247** (.106)	.061 (.099)
logworkdays	.366* (.114)	.376* (.114)			.370* (.113)	.390* (.114)
ddisaster	.250 (.166)	.179 (.317)	.392** (.175)	.262 (.332)	.264 (.164)	.057 (.316)
dcommunist	.210** (.087)	.129*** (.075)	.195** (.092)	.109 (.079)	.068 (.139)	-.023 (.125)
dpoorpeasant	-.043 (.075)	-.007 (.065)	-.040 (.079)	.001 (.068)	-.044 (.074)	-.008 (.064)
drichpeasant	.125 (.177)	.199 (.155)	.178 (.187)	.229 (.162)	.085 (.174)	.180 (.153)
dlandlord	-.025 (.160)	.107 (.147)	-.037 (.166)	.132 (.149)	-.062 (.158)	.102 (.145)
dcadre	-.010 (.108)	.076 (.096)	.240** (.103)	.267* (.089)	-.137 (.130)	.045 (.116)
dnewcadre	-.086 (.170)	-.136 (.151)	-.010 (.175)	-.082 (.152)	-.146 (.186)	-.085 (.162)
doldcadre	-.098 (.123)	.010 (.107)	.022 (.125)	.101 (.109)	-.065 (.123)	.041 (.107)
dwagejob	.230* (.086)	.153*** (.084)			.253* (.090)	.133 (.090)
dsideline	.183** (.085)	.214* (.082)			.058 (.096)	.073 (.093)
dcomm&cadre					.349 (.212)	.072 (.190)
dcomm&wagejob					-.151 (.204)	.019 (.180)
dcomm&sideline					.440** (.177)	.486* (.165)
village dummies	no	yes	no	yes	no	yes
d.f.	234	219	237	222	231	216
F stat.	6.41	8.16	4.74	7.21	6.21	7.98
adj. R^2	.289	.490	.191	.432	.312	.505

Notes:
1. Constant terms are not reported here.
2. * indicates significance at the 1% confidence level, ** at 5%, and *** at 10%.
3. Numbers in parentheses are standard errors.
4. All regressions weight observations by family size. To make the results representative, observations are also weighted by village population relative to village sample population. (The sample size in each village is fairly uniform, but village populations vary substantially, from 170 to 4,714 people.)

Table 5
Estimates of Per Capita Household Income,
Semi-Log Specification
(n=254)

	1	2	3	4	5	6
landpc	.342* (.063)	.231* (.073)	.299* (.067)	.214* (.076)	.343* (.063)	.232* (.072)
plots	.017 (.033)	.047 (.034)	.058*** (.034)	.069*** (.037)	.011 (.032)	.030 (.035)
hillarea	-1.190** (.545)	-.180 (.631)	-1.454** (.572)	-.207 (.661)	-1.407* (.544)	-.718 (.668)
pop	-.021 (.035)	-.036 (.032)	-.037 (.036)	-.041 (.033)	-.018 (.034)	.023 (.032)
adultspc	.539* (.199)	.573* (.176)	.306 (.204)	.410** (.178)	.487** (.198)	.510* (.175)
maleadult	-.356 (.297)	-.146 (.258)	-.283 (.319)	-.002 (.275)	-.267 (.297)	-.013 (.260)
age	.002 (.004)	-.005 (.003)	.002 (.004)	-.007*** (.004)	.003 (.004)	-.005 (.003)
educ	.042** (.020)	.009 (.019)	.047** (.021)	.005 (.020)	.043** (.020)	.011 (.019)
workdays	.002* (.000)	.002* (.000)			.002* (.000)	.002* (.000)
ddisaster	.317*** (.162)	.167 (.310)	.422** (.172)	.234 (.329)	.327** (.160)	.056 (.309)
dcommunist	.217** (.086)	.121 (.074)	.209** (.092)	.108 (.079)	.082 (.138)	.002 (.123)
dpoorpeasant	-.030 (.074)	.009 (.064)	-.031 (.079)	.011 (.068)	-.033 (.074)	.010 (.064)
drichpeasant	.180 (.175)	.226 (.152)	.218 (.186)	.244 (.162)	.139 (.173)	.209 (.151)
dlandlord	-.004 (.158)	.121 (.143)	-.013 (.165)	.148 (.147)	-.037 (.156)	.121 (.142)
dcadre	-.022 (.107)	.059 (.094)	.235** (.103)	.269* (.088)	-.156 (.128)	.045 (.113)
dnewcadre	-.056 (.169)	-.122 (.149)	-.021 (.174)	-.088 (.151)	-.133 (.186)	-.060 (.161)
[illegible]cadre	-.066 (.122)	.022 (.106)	.032 (.125)	.108 (.109)	-.040 (.122)	.048 (.106)
[illegible]gejob	.204** (.084)	.152*** (.082)			.233* (.088)	.145*** (.087)
dsideline	.157*** (.085)	.215* (.081)			.044 (.095)	.082 (.092)
dcomm&cadre					.372*** (.210)	.050 (.187)
dcomm&wagejob					-.159 (.202)	-.017 (.177)
dcomm&sideline					.405** (.176)	.456* (.163)
village dummies	no	yes	no	yes	no	yes
d.f.	234	219	237	222	231	216
F stat.	6.72	8.69	4.79	7.32	6.45	8.46
adj. R^2	.301	.508	.193	.437	.322	.522

Notes:
1. Constant terms are not reported here.
2. * indicates significance at the 1% confidence level, ** at 5%, and *** at 10%.
3. Numbers in parentheses are standard errors.
4. All regressions weight observations by family size. To make the results representative, observations are also weighted by village population relative to village sample population. (The sample size in each village is fairly uniform, but village populations vary substantially, from 170 to 4,714 people.)

higher than those of non-cadre families with wage employment. Past and/or present cadre status has no special advantages; however, access to wage employment does.[20]

Perhaps the most interesting results are those related to Communist Party membership. In regressions without interaction terms, the coefficient on Party membership is significant and positive. In these regressions the coefficient on the dummy variable indicating presence of a sideline business (dsideline) is also significant and positive. When interaction terms are included the coefficients on dcommunist are smaller (or negative) and no longer significant; dsideline is also no longer significant. The coefficient on the interaction term is, however, large and significant. These results imply that having a Party member or having a sideline business does not by itself increase a household's per capita income. Rather, households that have both Party members and sideline businesses enjoy higher incomes.[21]

The regression estimates thus provide evidence to support the view that some Party members have connections and knowledge that they use to advantage when running private businesses. The magnitude of this advantage is substantial. Calculations using coefficients from regression 6 in table 4 yield the following numbers: holding all other variables constant, having a sideline business raises household per capita income by 8 percent (compared to income without a sideline). Having a Party member in the household *lowers* per capita income by 2 percent. Having both a sideline *and* a Party member raises per capita income by 48 percent compared to that with a sideline but no Party member, and by 59 percent compared to that with neither a sideline nor a Party member. The joint effect of having both a sideline and a Party member on the level of household income per capita is thus

---

[20]Interestingly, in informal conversations village cadres often expressed the view that their work was time-consuming and not terribly remunerative.

[21]Note that in our sample (before weighting) 46 households had sideline businesses, 57 had a family member in the Communist Party, and 12 had both a sideline and a party member. After weighting, the numbers of households in these categories increases to 62, 64, and 19, respectively.

large.[22]

Inequality decomposition

Tables 6 and 7 contain the results of inequality decomposition for the Theil-L index using the regression estimates from regression 6 in table 4. Calculation of the different components of inequality follows the formulae set out in section II. Between-group inequality accounts for the largest share of overall inequality (77 percent). Within-inequality accounts for 43 percent of total inequality. This is less than the amount suggested by the regression R^2 of .58, but still substantial. The interaction term is negative and equal to 20 percent of total inequality.

Table 6
Bottom-Up Inequality Decomposition Using the Theil-L Index

Full Sample Theil-L: 0.169
Between Inequality: 0.130 (76.8%)
Within Inequality: 0.072 (42.7%)
Interaction Adjustment: -0.033 (-19.5%)

NOTE: Calculations are for the entire sample, composed of 1028 individuals in 254 households, weighted to correct for sampling bias across villages. Numbers in parentheses are percentages of aggregate inequality.

Table 7 disaggregates between-group inequality to show the individual contributions of different attributes. Of the various attributes, land per capita (landpc) has the largest contribution to total inequality. Other attributes that contribute substantially are days worked per adult (workdays) and the interaction between Party membership and household sideline (dcomm&sideline). To the extent that families decide how much time to work voluntarily, the fact that days worked per worker contributes substantially to inequality is perhaps unobjectionable. Days worked, however, can be influenced by factors beyond the household's control such as illness or disability. Since the data set does not contain information on illness or related variables, we unfortunately cannot assess the

[22]Per capita income with both a party member and a sideline is 63 percent higher than with a party member but no sideline. These percentage increases are calculated using the geometric means of predicted household incomes with and without the relevant attributes.

Table 7
Between Inequality: Contributions of Individual Attributes

Variable	Coefficient	Theil-L	Contribution to Overall Inequality	
			Absolute	Share
landpc	.497*	.062	.0306	.181
plots	.096	.050	.0048	.028
hillarea	-.712	.003	-.0023	-.013
pop	-.113	.037	-.0041	-.024
adultspc	.489*	.017	.0082	.048
maleadult	-.008	.006	-.0001	-.000
age	-.167	.022	-.0037	-.022
educ	.061	.053	.0032	.019
workdays	.390*	.053	.0208	.123
ddisastr	.057	.035	.0019	.012
dcommunist	-.023	.107	.0025	-.015
dpoorpeasant	-.008	.105	.0008	-.005
drichpeasant	.180	.024	.0043	.026
dlandlord	.102	.033	.0034	.020
dcadre	.045	.081	.0037	.022
dnewcadre	-.085	.027	-.0023	-.014
doldcadre	.041	.053	.0022	.013
dwagejob	.133	.117	.0156	.093
dsideline	.073	.106	.0078	.046
dcomm&cadre	.072	.062	.0045	.027
dcomm&wagejob	.019	.081	.0015	.009
dcomm&sideline	.486*	.045	.0218	.129
village dummies		.036 (average)	.0110	.065

Notes:
1. * indicates significance at the 1% confidence level, ** at 5%, and *** at 10%.
2. Theil-L indices are calculated using the unlogged values of variables that appear in logged form, and using the exponentiated value of variables that appear in unlogged form, in the regression equation. That is, the individual Theil-L indices are calculated in accordance with the underlying income equation as shown in expression (5).
3. As elsewhere, all values are weighted to correct for sampling bias across villages.

importance of voluntary versus involuntary variations in days worked.

The interaction between Party membership and household sideline (dcomm&sideline) contributes 13 percent of total inequality. Other status variables are relatively unimportant. The contribution of the wage job dummy (dwagejob) is relatively large, which highlights the importance of uneven access to wage employment. The contribution of village of residence to inequality is only 7 percent, considerably less than in the top-down decomposition.

As mentioned above, an attribute's contribution to inequality depends both on the magnitude of its coefficient and on the degree of inequality in its distribution. Land per capita, adults per capita, days worked per adult, and the Party/sideline interaction term all have large coefficients (table 7). Consequently, even though they are distributed fairly evenly across the population, they contribute noticeably to income inequality. Attributes that are (relatively) unequally distributed include Party membership and poor peasant background. These variables, however, have small coefficients and so their contributions to income inequality are fairly low.

IV. Conclusion

In this paper we outline an alternative approach to inequality decomposition and use this approach to analyze inequality in rural China. Our approach begins with regression analysis. The regression is specified so as to capture the underlying relationship between income and its determinants, with corrections if appropriate for simultaneity and any other specification problems. The regression results are then used in the inequality decomposition.

The mechanics of the decomposition will depend on both the functional form of the income equation and the chosen inequality index. We describe the general method, and go through the mechanics for the case where the regression specification is log-linear and inequality is measured using the Theil-L index. Within and between components of inequality are calculated using estimates of conditional means from the regression.

The decomposition can take one of two paths. If the analyst is committed to a particular group division, for example, regional groups, then the decomposition can simply be calculated over

these groups. Inequality between and within regional groups, however, can now be broken down further into their systematic and non-systematic components. If the analyst is agnostic about group divisions, or if variables over which one wants to decompose are continuous, then it is natural to let the income equation determine the structure of the decomposition. In this case groups are defined as sets of individuals who are identical in all attributes. Between inequality is now interpreted as the systematic component of inequality, that is, as the share of inequality that arises because of variation in location, education, ethnic group, and other determinants of income. Within inequality is that component of inequality that is non-systematic or "residual" after holding constant all the factors that determine income.

From a policy perspective, reinterpreting the meaning of between- and within-inequality has certain advantages. When using the standard approach, the analyst must somewhat arbitrarily choose one or two variables over which to calculate the decomposition. In contrast, the bottom-up approach allows one to discover which variables are most important in explaining inequality. Indeed, by revealing which variables are most important, this approach provides information useful for targeting policies to reduce inequality. For example, if the decomposition reveals that inequality in education explains a large share of overall inequality, then government efforts to reduce inequality can focus on changing the distribution of education. If the decomposition reveals that location is most important, then distributional policies can target poor localities.

Of course, the decomposition might reveal that within inequality exceeds between inequality, in which case the majority of inequality is non-systematic. In this event the most effective policy may simply be to give transfers directly to the poor. In other words, the greater the contribution of within inequality, the stronger the argument for direct redistribution, while the greater is the role of between inequality, the stronger the argument for targeted or project specific interventions.[23]

An additional advantage for policy of using the bottom-up approach is that it can yield

---

[23]The appropriate approach to reducing inequality depends on both the benefits and costs of different policies. Bottom-up decomposition provides information only about potential benefits.

predictions of how different policies would change inequality. Overall inequality is treated as a function of inequality in the variables that determine income. Specifically, inequality contains the weighted average of inequality in these variables, where the weights are given by estimated coefficients from the regression. Thus calculating how a change in the distribution of one or more independent variables changes total inequality is relatively straightforward. For example, one could estimate how much inequality would decline if the government adopted policies that ensured universal primary education.

Our analysis of inequality for a sample of households in rural China illustrates some of these features. The regression results identify key determinants of income in this part of China. Land per capita, adults per capita, days worked per adult, and village of residence all have significant, and in some cases large, effects on the level of per capita income. Certain types of elite status or special access are also important. For example, presence of a wage earner is significantly and positively associated with income, which suggests that those households with access to wage employment are at an advantage. Communist Party membership has a significant and positive impact on income, but only when in combination with a household sideline. Class background and cadre status do not appear to be significant. These results suggest that the debate over who benefits during the reform of a socialist economy does not have a clear winner: in China some of the usual household characteristics matter, and some status variables matter.

Turning to the question of income distribution, we decompose inequality for our sample using both the standard approach and the bottom-up approach. Standard decomposition over groups with and without Communist Party members finds that about 4 percent of total inequality is between groups, and the rest is within groups. Decomposition by village of residence shows that roughly 40 percent of total inequality is between villages and 60 percent is within villages.

The bottom-up decomposition yields a different and more informative picture. Holding other variables constant, Party membership by itself explains only a small fraction of inequality; in interaction with other attributes (presence of a sideline, wage job, and cadre), it explains about 15

percent of inequality. Village effects explain 7 percent of inequality. The differences between these results and those of the top-down decomposition occur in part because omitted variables are held constant, and in part because the top-down approach implicitly assumes the income equation is linear. In addition, the bottom-up decomposition reveals which of the many determinants of income are most important in explaining inequality. For our sample, those variables are land per capita, days worked per adult, and the interaction between Party membership and household sideline.

These findings yield some useful implications for distributional policy in Zouping County and similar parts of China. One implication is that policy makers may want to measures that ensure or maintain equal access to land can have a positive distributional impact. Even though land per capita is distributed relatively equally, land has a large effect on average income. Thus a small increase in the inequality of land distribution could have a substantial effect on overall inequality. Similarly, small changes in inequality of days worked per adult can have a large effect on inequality. Consequently, efforts to aid households hindered by illness or disability may be desirable.

Reducing the special advantages of Party members in establishing and operating private businesses also has the potential to reduce inequality. The measures appropriate to accomplish this goal will depend on the factors that give rise to these advantages. Lastly, policy makers should not overlook non-systematic factors. The contribution of within inequality is relatively large–indeed, it exceeds that of any single component of between inequality. Consequently direct redistribution may be a preferred policy even if the cost of measuring household incomes is higher than that of identifying households with specific attributes.

BIBLIOGRAPHY

Alesina, Alberto, and Dani Rodrik. 1994. Distributive politics and economic growth. *Quarterly Journal of Economics* 109(2): 465-490. May.

Anand, Sudhir. 1983. *Inequality and Poverty in Malaysia: Measurement and Decomposition.* New York: Oxford University Press.

Anand, Sudhir, and Ravi Kanbur. 1993. Inequality and development: A critique. *Journal of Development Economics* 41(1): 19-43. June.

Atkinson, Anthony. 1970. Measurement of inequality. *Journal of Economic Theory* 2(3): 244-263. September.

Fei, John C. H., Gustav Ranis, and Shirley W. Y. Kuo. 1978. Growth and the Family Distribution of Income by Factor Components. *Quarterly Journal of Economics* 92(1): 17-53. February.

Fields, Gary S. 1980. *Poverty, Inequality and Development.* New York: Cambridge University Press.

Fields, Gary S., and George H. Jakubson. 1992. New evidence on the Kuznets curve. Paper presented at the New England Universities Development Consortium Conference, Boston University.

Griffin, Keith, and Renwei Zhao, eds. 1993. *The Distribution of Income in China.* New York: The MacMillan Press.

Hare, Denise. 1994. Rural nonagricultural activities and their impact on the distribution of income: Evidence from farm households in Southern China. *China Economic Review* 4(1): 59-82. Spring.

Hussain, Athar, Peter Lanjouw, and Nicholas Stern. 1991. Income inequalities in China: Evidence from household survey data. London School of Economics Development Economics Research Programme working paper CP No. 18. November.

Knight, John, and Lina Song. 1993a. The spatial contribution to income inequality in rural China. *Cambridge Journal of Economics* 17(2): 195-213. June.

Knight, John, and Lina Song. 1993b. Workers in China's rural industries. Chapter 6 in Keith Griffin

and Renwei Zhao, eds. 1993. *The Distribution of Income in China.* New York: The MacMillan Press. pp. 173-215.

Khan, Azizur Rahman, Keith Griffin, Carl Riskin, and Renwei Zhao. 1992. Household income and its distribution in China. *China Quarterly* 132: 1029-1061. December.

Kuznets, Simon. 1955. Economic growth and income inequality. *American Economic Review*: 1-28. March.

McAuley, Alastair. 1991. The economic transition in Eastern Europe: Employment, income distribution, and the social security net. *Oxford Review of Economic Policy* 7(4): 93-105. Winter.

Nee, Victor. 1989. A theory of market transition: From redistribution to markets in state socialism. *American Sociological Review* 54(5): 663-81.

Nee, Victor. 1991. Social inequalities in reforming state socialism: Between redistribution and markets in China. *American Sociological Review* 56(3): 267-82. June.

Oi, Jean C. 1989. *State and Peasant in Contemporary China.* Berkeley: University of California Press.

Putterman, Louis. 1993. *Continuity and Change in China's Rural Development: Collective and Reform Eras in Perspective.* New York: Oxford University Press.

Oi, Jean C. 1992. Fiscal reform and the economic foundations of local state corporatism in China. *World Politics* 45(1): 99-126. October.

Riskin, Carl. 1993. Income distribution and poverty in rural China. Chapter 5 in Keith Griffin and Renwei Zhao, eds. 1993. *The Distribution of Income in China.* New York: The MacMillan Press. pp. 135-170.

Riskin, Carl. 1994. Chinese rural poverty: Marginalized or dispersed? *American Economic Review* 84(2): 281-284. May.

Rozelle, Scott. 1993. Stagnation without equity: Changing patterns of income and inequality in China's post-reform rural economy. Unpublished manuscript.

Shirk, Susan. 1989. The political economy of Chinese industrial reform. In Victor Nee and

David Stark, eds. *Remaking the Economic Institutions of Socialism.* Stanford: Stanford University Press. pp. 328-364.

Shorrocks, A. F. 1980. The class of additively decomposable inequality measures. *Econometrica* 48(3): 613-625. April.

Shorrocks, A. F. 1982. Inequality decomposition by factor components. *Econometrica* 50(1): 193-211. January.

Shorrocks, A. F. 1983. The Impact of Income Components on the Distribution of Family Incomes. *Quarterly Journal of Economics* 98(2): 311-326. May.

Sicular, Terry. 1993. Redefining plan, market and the state: Lessons from an investigation of rural commerce in Zouping County. Unpublished manuscript.

Szelényi, Iván. 1986-87. The prospects and limits of the East European new class project: An auto-critical reflection on the intellectuals on the road to class power. *Politics and Society* 15(2): 103-44. September.

Szelényi, Iván. 1988. *Socialist Entrepreneurs. Embourgeoisement in Rural Hungary.* Madison: Wisconsin University Press.

Szelenyi, Szonja. 1987. Social inequality and party membership: Patterns of recruitment into the Hungarian Socialist Workers' Party. *American Sociological Review* 52(5): 559-573. October.

Róna-Tas, Akos, and Tamás Kolosi. 1993. The first shall be last? The social consequences of the transition from socialism. Unpublished manuscript.

The World Bank. 1992. China: Strategies for reducing poverty in the 1990s. World Bank Country Study CIP 92-34227.

Appendix Table 1

Economic Profile of Zouping County, 1990

Population	670,571
Household size (national average)	3.96 (3.97)
Cultivated land per capita (national average)	0.11 ha (0.08 ha)
GDP per capita (national)	1,280 yuan (1,558 yuan)
Share of agriculture in GDP (national share)	54% (28%)
Share of industry and construction in GDP (national share)	25% (44%)
Per capita net income of rural households (national average)	702 yuan (630 yuan)
Annual growth in net material product, 1980-1990, constant prices (national rate)	10.4% (8.7%)

Sources: Interviews; State Statistical Bureau (1991).

Note: Household income statistics in this table are the official county and national averages. They are not strictly comparable to income statistics used in this paper, in part because the method of calculation is different. For example, in the official statistics retained output is valued at planned prices, while we value retained output at market prices.

Appendix Table 2
Variable Names and Definitions

Variable	Description
income[1]	household net income (yuan)
incomepc	household net income per capita (income/pop)
land	cultivated land area (mu)[2]
landpc	cultivated land per capita (land/pop)
plots	number of plots of cultivated land
hillarea	share of cultivated land that is not flat (hilly or mountainous)
pop	number of people resident in the household for one or more months
adults	number of working-age adults in the household (ages 18-65)
adultspc	number of adults per person (adults/pop)
maleadult	share of adults that are male
age	average age of adults
educ	average years education of adults
workdays	days worked per adult (standardized 8 hour days)
ddisaster	=1 if in 1990 the household experienced natural disaster that affected cultivation; if not, =0
dcommunist	=1 if the household has a Communist Party member =1; if not, =0
dpoorpeasant	=1 if class background is landless or poor peasant; if not, =0
drichpeasant	=1 if class background is rich peasant; if not, =0
dlandlord	=1 if class background is landlord; if not, =0
dcadre[3]	=1 if the household had a cadre in 1990 and before decollectivization; otherwise, =0
dnewcadre	=1 if the household had a cadre in 1990 but not before decollectivization; otherwise =0
doldcadre	=1 if the household had a cadre before decollectivization, but not in 1990; otherwise =0
dwagejob	=1 if a household member has a wage employment; otherwise =0
dsideline	=1 if the household has a (nonagricultural) sideline activity; otherwise =0
dcomm&cadre	=1 if the household has a Party member and a cadre in 1990; otherwise =0
dcomm&wagejob	=1 if the household has a Party member and a wage earner; otherwise =0
dcomm&sideline	=1 if the household has a Party member and a sideline; otherwise =0

Notes:
1. Retained output is valued at market prices.
2. Fifteen *mu* equal one hectare.
3. Almost all households with a cadre in 1990 also reported having a cadre before 1990 but after decollectivization. So households with cadres in 1990 effectively overlap with households that have had cadres during the reform period.

Appendix Table 3
Estimates of Household Income,
Log-Linear Specification
(n=254)

	1	2	3	4	5	6
logland	.669* (.121)	.498* (.179)	.566* (.126)	.456** (.186)	.662* (.123)	.497* (.179)
logplots	-.046 (.125)	.167 (.135)	.135 (.129)	.255*** (.141)	-.060 (.124)	.096 (.135)
hillarea	-.780 (.579)	-.106 (.644)	-1.125*** (.603)	-.148 (.663)	-1.018*** (.582)	-.712 (.682)
logpop	.286*** (.169)	.335 (.205)	.304*** (.176)	.363*** (.213)	.294*** (.167)	.391*** (.203)
adultspc	.532* (.205)	.547* (.182)	.331 (.208)	.416** (.182)	.480** (.203)	.489* (.182)
maleadult	-.341 (.300)	-.141 (.263)	-.291 (.319)	-.025 (.276)	-.239 (.300)	-.008 (.265)
logage	.122 (.164)	-.169 (.147)	.095 (.172)	-.257*** (.153)	.145 (.162)	-.167 (.146)
logeduc	.235** (.107)	.052 (.099)	.282** (.113)	.037 (.104)	.247** (.106)	.061 (.099)
logworkdays	.366* (.114)	.376* (.114)			.370* (.113)	.390* (.114)
ddisaster	.250 (.166)	.179 (.317)	.392** (.175)	.262 (.332)	.264 (.164)	.057 (.316)
dcommunist	.210** (.087)	.129*** (.075)	.195** (.092)	.109 (.079)	.068 (.139)	-.023 (.125)
dpoorpeasant	-.043 (.075)	-.007 (.065)	-.040 (.079)	.001 (.068)	-.044 (.074)	-.008 (.064)
drichpeasant	.125 (.177)	.199 (.155)	.178 (.187)	.229 (.162)	.085 (.174)	.180 (.153)
dlandlord	-.025 (.160)	.107 (.147)	-.037 (.166)	.132 (.149)	-.062 (.158)	.102 (.145)
dcadre	-.010 (.108)	.076 (.096)	.240** (.103)	.267* (.089)	-.137 (.130)	.045 (.116)
dnewcadre	-.086 (.170)	-.136 (.151)	-.010 (.175)	-.082 (.152)	-.146 (.186)	-.085 (.162)
doldcadre	-.098 (.123)	.010 (.107)	.022 (.125)	.101 (.109)	-.065 (.123)	.041 (.107)
dwagejob	.230* (.086)	.153*** (.084)			.253* (.090)	.133 (.090)
dsideline	.183** (.085)	.214* (.082)			.058 (.096)	.073 (.093)
dcomm&cadre					.349 (.212)	.072 (.190)
dcomm&wagejob					-.151 (.204)	.019 (.180)
dcomm&sideline					.440** (.177)	.486* (.165)
village dummies	no	yes	no	yes	no	yes
d.f.	234	219	237	222	231	216
F stat.	9.70	10.72	8.18	9.73	9.14	10.41
adj. R^2	.395	.567	.312	.517	.415	.579

Notes:
1. Constant terms are not reported here.
2. * indicates significance at the 1% confidence level, ** at 5%, and *** at 10%.
3. Numbers in parentheses are standard errors.
4. All regressions weight observations by family size. To make the results representative, observations are also weighted by village population relative to village sample population. (The sample size in each village is fairly uniform, but village populations vary substantially, from 170 to 4,714 people.)

Appendix Table 4
Estimates of Household Income,
Semi-Log Specification
(n=254)

	1	2	3	4	5	6
land	.093* (.017)	.057* (.022)	.079* (.018)	.050** (.023)	.095* (.017)	.058* (.022)
plots	.019 (.033)	.063*** (.035)	.064*** (.035)	.086** (.037)	.011 (.033)	.045 (.035)
hillarea	-1.022*** (.562)	-.058 (.641)	-1.328** (.594)	-.107 (.672)	-1.234** (.560)	-.615 (.678)
pop	-.074*** (.038)	.111* (.042)	.078*** (.040)	.116* (.043)	.075** (.038)	.124* (.041)
adultspc	.470** (.201)	.508* (.177)	.237 (.206)	.339*** (.180)	.410** (.200)	.440** (.176)
maleadult	-.356 (.300)	-.148 (.260)	-.288 (.322)	-.002 (.278)	-.250 (.300)	-.005 (.262)
age	.000 (.004)	-.007*** (.004)	-.000 (.004)	-.009** (.004)	.001 (.004)	-.007** (.003)
educ	.045** (.020)	.008 (.019)	.050** (.022)	.004 (.020)	.047** (.020)	.009 (.019)
workdays	.002* (.000)	.002* (.000)			.002* (.000)	.002* (.000)
ddisaster	.277*** (.164)	.185 (.313)	.391** (.174)	.249 (.333)	.287*** (.162)	.066 (.311)
dcommunist	.213** (.087)	.123*** (.075)	.207** (.093)	.111 (.080)	.059 (.139)	-.007 (.124)
dpoorpeasant	-.057 (.075)	-.017 (.064)	-.058 (.080)	-.014 (.068)	-.060 (.074)	-.016 (.064)
drichpeasant	.160 (.177)	.215 (.154)	.200 (.188)	.232 (.164)	.117 (.174)	.199 (.152)
dlandlord	.014 (.160)	.109 (.145)	-.008 (.167)	.137 (.149)	-.021 (.158)	.110 (.143)
dcadre	-.020 (.108)	.064 (.095)	.243** (.104)	.280* (.090)	-.148 (.129)	.056 (.115)
dnewcadre	-.060 (.171)	-.114 (.151)	-.019 (.176)	-.079 (.153)	-.123 (.187)	-.042 (.163)
doldcadre	-.067 (.123)	.016 (.107)	.035 (.127)	.102 (.111)	-.036 (.123)	.044 (.106)
dwagejob	.213** (.085)	.155*** (.084)			.242* (.089)	.146 (.089)
dsideline	.177** (.086)	.216* (.082)			.049 (.097)	.076 (.093)
dcomm&cadre					.347 (.212)	.028 (.189)
dcomm&wagejob					-.141 (.204)	.002 (.179)
dcomm&sideline					.462* (.178)	.487* (.165)
village dummies	no	yes	no	yes	no	yes
d.f.	234	219	237	222	231	216
F stat.	9.59	10.98	7.70	9.46	9.09	10.68
adj. R^2	.392	.573	.298	.509	.413	.586

Notes:
1. Constant terms are not reported here.
2. * indicates significance at the 1% confidence level, ** at 5%, and *** at 10%.
3. Numbers in parentheses are standard errors.
4. All regressions weight observations by family size. To make the results representative, observations are also weighted by village population relative to village sample population. (The sample size in each village is fairly uniform, but village populations vary substantially, from 170 to 4,714 people.)

Appendix Table 5
Summary Statistics for Regression Variables, Unweighted

variable	mean	standard deviation	minimum	maximum
income	4756	3475	690	26262
logincome	8.260	.642	6.537	10.176
incomepc	1173	734	154	5252
logincomepc	6.907	.561	5.035	8.566
land	6.9	2.7	1.9	17.0
logland	1.847	.430	.642	2.833
landpc	1.72	.55	.67	5.67
loglandpc	.495	.315	-.405	1.735
plots	3.6	1.3	1	9
logplots	1.198	.392	0	2.197
hillarea	.029	.096	0	.54
pop	4.0	1.1	1	7
logpop	1.353	.321	0	1.946
adultspc	.66	.19	.33	1
maleadult	.52	.13	.25	1
age	37.4	8.29	20	72
educ	5.4	1.8	1.5	11
workdays	244.4	79.5	50	548.9
logworkdays	5.444	.339	3.912	6.308
ddisaster	.14	.35	0	1
dcommunist	.22	.42	0	1
dpoorpeasant	.59	.49	0	1
drichpeasant	.04	.19	0	1
dlandlord	.06	.23	0	1
dcadre	.13	.33	0	1
dnewcadre	.06	.24	0	1
doldcadre	.07	.26	0	1
dwagejob	.48	.50	0	1
dsideline	.18	.39	0	1
dcomm&cadre	.13	.33	0	1
dcomm&wagejob	.15	.36	0	1
dcomm&sideline	.05	.21	0	1

Appendix Table 6
Summary Statistics for Regression Variables, Weighted

variable	mean	standard deviation	minimum	maximum
income	4832	3356	690	26262
logincome	6.848	.582	5.035	8.566
incomepc	1116	716	154	5252
logincomepc	6.848	.581	5.035	8.566
land	6.99	2.71	1.9	17.0
logland	1.864	.419	.642	2.833
landpc	1.63	.61	.67	5.67
loglandpc	.424	.349	-.405	1.735
plots	3.8	1.2	1	9
logplots	1.289	.328	0	2.197
hillarea	.027	.077	0	.54
pop	4.4	1.1	1	7
logpop	1.439	.279	0	1.946
adultspc	.64	.18	.33	1
maleadult	.51	.11	.25	1
age	37.4	8.28	20	72
educ	5.6	1.7	1.5	11
workdays	259.2	82.03	50	548.9
logworkdays	5.504	.337	3.912	6.308
ddisaster	.05	.23	0	1
dcommunist	.25	.43	0	1
dpoorpeasant	.62	.49	0	1
drichpeasant	.04	.19	0	1
dlandlord	.05	.22	0	1
dcadre	.15	.36	0	1
dnewcadre	.04	.20	0	1
doldcadre	.09	.28	0	1
dwagejob	.53	.50	0	1
dsideline	.24	.43	0	1
dcomm&cadre	.11	.31	0	1
dcomm&wagejob	.16	.36	0	1
dcomm&sideline	.07	.26	0	1

Note: Observations are weighted by household population and to correct for sampling bias among villages.