The Dynamics of Inequality and Growth in Rural China: Does Higher Inequality Impede Growth?

Dwayne Benjamin
Department of Economics
University of Toronto

Loren Brandt
Department of Economics
University of Toronto

John Giles
Department of Economics
Michigan State University

Preliminary and Incomplete

This draft: April 9th, 2004[*]

[*]Benjamin and Brandt thank the Social Sciences and Humanities Research Council of Canada for financial support. Giles gratefully acknowledges financial support from the National Science Foundation and the Michigan State Intramural Research Grants Program.

The dynamics of inequality and growth in rural China:
Does higher inequality impede growth?

Abstract

This paper explores the empirical linkages between income inequality and growth in rural China in the post-reform period. Since the early 1980's, China experienced high rates of growth accompanied by increases of income inequality. As long as living standards rose for everyone, widening income gaps were viewed as the inevitable, temporary consequence of the transition process. However, there is now concern that recent increases of inequality threaten future growth. Our primary objective is to evaluate whether there is any evidence from recent experience that confirms inequality can hinder growth. The analysis is based on a large-scale, detailed household survey from over 100 villages, spanning the period 1986 to 1999. We distill the household-level data to the village-level, creating a panel of 100 villages for this time period. Taking the village as the unit of observation, we estimate models relating a village's growth rate to its initial level of inequality, and a set of covariates. Our starting point is a dynamic panel-data specification, similar in spirit to recent country-level studies in the macro and development literature (e.g., Forbes 2000; Lundberg and Squire, 2003). Within that framework, we find no evidence suggesting that inequality reduces growth, and this conclusion is robust to the consideration of a variety of econometric issues. However, we argue that the potentially long run relationship between inequality and growth is better detected in a cross-section framework, as in Easterly (2002) and Persson and Tabellini (1994). One reason for this is that the higher frequency data are dominated by measurement and (possibly) short-run positive co-movements of inequality and growth, consistent with the "Kuznets" process of inequality accompanying growth. Within the cross-section framework, we find evidence that levels of inequality in 1986 are negatively related to the growth of village incomes through 1999, suggesting that higher inequality can indeed hurt growth in the long run. While we do not yet identify the specific institutional channels, we do find strong evidence that inequality adversely affects development of the non-agricultural sector, and it is this stunted growth that underlies the broader negative impact of inequality on growth of household incomes.

1.0 Introduction

Since the beginning of economic reforms, rural Chinese incomes were characterized by rising average levels, accompanied by increasing dispersion or inequality. For the most part, the widening disparities were not viewed as a serious problem, at least as long as incomes were rising steadily, and everyone was better off. More recently however, alarms have been sounded, both in response to continued widening of inequality and worsening absolute conditions for the poorest. For example, a recent World Bank country report highlights widening income disparities as one of the most serious threats to China's future prosperity.[1] One concern is that regional inequality may prompt inter-regional redistributive policies that may harm growth. However, there are also widening disparities within regions (localities), and political pressure may threaten the long run viability of economic reforms. Furthermore, drawing on results from the cross-country growth literature, higher inequality may slow down growth for reasons beyond immediate political pressure.

Our objective in this paper is to use a rich household data set covering households from over 100 villages from 1986 through 1999, to explore linkages between local inequality and the growth of household incomes, in order to evaluate whether there is any evidence in the Chinese context that inequality reduces growth. In essence, this entails estimating village-level regressions that mimic those in the cross-country growth literature. As it turns out, there is considerable overlap in the issues that we must consider and those confronted in the cross-country literature. While we are able to draw on existing insights, the Chinese village-level data have their own peculiarities, and advantages over country-level data – most notably that the key variables are constructed in a consistent manner – and we are able to address some of the methodological questions raised in the cross-country studies.

We begin with a brief overview of the conceptual and empirical literature concerning inequality and growth, especially highlighting some of the econometric and interpretation issues. After describing the basic patterns in the household data, and justifying the analysis at the village level, we turn to our

[1] See World Bank, 2003.

main empirical work. Our base specification is similar to that employed by Forbes (2000), a panel-data regression linking village growth to previous inequality and a set of control variables. We employ a variety of estimation strategies, paying particular attention to the possibility that the empirical relationship might be dominated by measurement error, or other mechanical linkages between measured inequality and growth. We also try to account for the possibility that the main equation is contaminated by a "Kuznets"-like relationship running from growth to inequality. Even employing a varied set of instruments, we find no evidence of a relationship between growth and inequality, in either casual direction.

As previous researchers have suggested, the high frequency variation in the panel data may not be the ideal framework in which to detect the possibly long run relationship between inequality and growth. We thus move to a cross-section study of growth, linking broad summaries of village economic performance over the 1986-1999 period to initial levels of inequality and other factors. Within this framework, we find strong evidence that inequality reduces growth. Furthermore, we find that one channel through which inequality might have mattered is in tilting villages away from development of their non-agricultural sectors. In the final section of the paper we offer some tentative interpretations of these findings.

2.0 Previous Studies

2.1 Conceptual background and cross-country evidence

There is an extensive literature examining the relationship between inequality and economic growth. This work identifies a number of alternative links between the two variables, and raises important issues about the causal nature of the relationship that has empirical implications. Kaldor (1956, 1967) and Lewis (1954) both suggest a positive link running between inequality and growth through the positive effect of inequality on savings. The latter arises because of differences in savings rates between the rich and the poor. The relationship is also causal, and suggests that increases in inequality can be good for growth. Kuznets (1955), on the other hand, was more interested in the link between growth and inequality, but of a non-causal nature. Kuznets saw the development process as entailing the reallocation of labor out of a low productivity sector (agriculture) and into the more productive non-agricultural sector. Differences in the rate of return to factors between sectors imply that an increase in inequality will "mechanically" accompany economic growth. However, this positive link need not be permanent, and at some point the relationship can turn negative, thus the hypothesized "inverted-U" in the relationship between inequality and growth.

More recent work suggests the possibility of a negative causal link, which can occur through a number of alternative channels. One possibility is that inequality influences growth through the political process. In models along the lines of Alessina and Rodrik (1994) and Persson and Tabellini (1994), high levels of inequality lead to redistributive pressures in the economy, largely through the fiscal system. This, in turn, is harmful for growth through its adverse effect on incentives for capital accumulation and investment in the economy. In the more heuristic model of Engerman and Sokoloff (2002), high levels of inequality are associated with a number of things that are detrimental for growth including redistributive policies, bad and inhibiting institutions, as well low levels of public goods' provision, such as education.

A second important link is through the interaction between wealth inequality (which we expect to be highly correlated with income inequality), and credit market imperfections. Because of a host of agency problems in both formal and informal credit markets in low-income countries, a household's ability to undertake new investments in human and physical capital will largely be a function of its own self-financing capability. The latter, in turn, will be highly correlated with the household's own wealth and incomes. In other words, all else equal, wealthier households will be able to invest more. Because investments in either physical or human capital are often not easily divisible, the overall level of investment in the economy should be negatively correlated with wealth inequality. Moreover, we might expect the effect of wealth inequality on investment to be more binding at lower levels of income (or wealth) in the economy (Banerjee and Duflo, 2003). Finally, differences in investment across households would tend to perpetuate and possibly increase inequality over time.

In principle, then, the causality between inequality and growth can be either negative or positive. Trying to identify the sign of the relationship will be confounded by the possibility of a positive correlation running between growth and inequality of the sort that Kuznets first described. As Lundberg and Squire point out, there is the possibility that growth and inequality are in fact being determined simultaneously.

The examination of the empirical relationship between inequality and growth has been severely handicapped by data limitations. The compilation by Deninger and Squire (1996) of a comprehensive and reasonably consistent data set on inequality for the period between 1960-1995 helped to relax some of these constraints, and contributed to an increase in the empirical work being done the last few years. However, key issues remain including the direction of causality; the functional form of the relationship; and the problems posed by measurement error and omitted variable bias for the identification of the effect of inequality on growth.

Much of the earlier literature was concerned with looking at the link between inequality and growth in a cross-sectional, reduced-form setting, with the causality running from inequality to growth. This relationship was estimated by adding a measure of inequality to a standard cross-country growth regression. In general, a negative relationship was found using a variety of data sets covering different periods and sets of countries.[2] For a relatively small sample of 43 countries, Persson and Tabellini (1994) also found that investment was negatively related to inequality, and attributed it to the effect of distributional conflict on tax policy and investment incentives. More recently, Easterly (2002) investigated the indirect effect of inequality on income levels through its effect on institutions and economic-openess (which he links directly with income levels).

One potential problem with the cross-sectional growth regressions is the potential correlation between initial inequality and country-level unobservables. One solution to the problem has been to take advantage of the Deninger and Squire data and look at the relationship in the context of panel data models, where fixed effects may be employed to control for unobserved differences across countries. Li

[2] Benabou (1996) and Perotti (1996) both provide summaries of some of empirical work up through 1995 or so.

and Zhou (1998) and Forbes (2000) both do so, and find that the effect of inequality on growth appears to be positive. Barro (2000) also uses panel data, but presents only the random effect estimates. He argues that most of the information on the relationship is coming from the cross-sectional dimension, and believes that measurement error in inequality poses serious problems for the fixed effect estimates.

Recently, Forbes has been criticized for ignoring the non-linearity of the relationship. Banerjee and Duflo (2003) argue on theoretical grounds that the relationship is nonlinear, and possibly quadratic. Using the same data as Forbes, they regress growth on lagged changes in inequality and the square of changes in inequality, and find that large changes in inequality in either direction are associated with larger future reductions in economic growth. They take this as support for the political economy models, which link growth to distributional politics.

Finally, as re-emphasized by Lundberg and Squire (2003), and other earlier papers, e.g. Anand and Kanbur (1993), there may be causality running from growth to inequality. This resultant correlation might contaminate the estimate of the effect of inequality on growth. A potential solution to the problem is to estimate a structural model of growth and inequality simultaneously, as do Lundberg and Squire. Using the Deininger and Squire data, they find that higher growth leads to greater inequality, and that greater inequality leads to higher growth. However, identification in the simultaneous equations framework (i.e., finding instruments) remains a major issue.

2.2 The China Angle

There is an extensive literature in China on the behavior of growth and inequality during reform. By all indications, increases in inequality have accompanied the rapid growth in China over the last two decades.[3] As long as incomes of all groups were rising, this increase has been seen as a natural corollary of the growth process. The possibility that inequality at any level may have explicit implications for future growth has generally not been part of the debate.

[3] See Benjamin, Brandt, and Giles (2003) for an extensive discussion and summary of the existing literature concerning inequality in China, as well as recent evidence on the trajectory of inequality.

More recently, however, there have been concerns that the increase in inequality is accelerating, largely a product of growing income differences between regions, as well as a widening gap in incomes between urban and rural households. Simultaneously, there have been calls for a host of policies, some regionally re-distributive, to redress the emerging inequality (see Wang 1999, for example). China's ambitious investment program to develop its western region can be seen in this light. In the growth literature, such policies may be seen as detrimental to future growth through their adverse incentive effects. In the China context, one interpretation for such behavior is that it is motivated by a larger fear that rising economic and social inequality will undermine the single basis for the political legitimacy of the current regime, namely, economic growth.

In this paper, we argue that a link between inequality and growth operates at a more than just the national level. Largely because of the high degree of administrative decentralization in China, but in the context of one-party rule, we observe considerable heterogeneity at the local level in the economic and political institutions that emerged with economic reform in China in the early 1980s. In the countryside, these factors influenced how resources were distributed among households with the break-up of the communes, and likely helped to shape local policy. Restrictions on factor mobility, like the household registration system, for much of the first decade of reform likely enhanced the effect of these policies on the trajectory of growth. The experience of rural China would appear to provide an excellent venue through which to look for links between inequality and growth, and to examine how a number of factors identified in the literature (e.g. credit market and political institutions) helped to influence the overall effect of inequality on growth.

3.0 Data

3.1 The RCRE Household Level Data

Our analysis will be at the village-level, but we construct most of the village measures from underlying household-level data. The data we use in this paper are the product of annual household surveys conducted by the Survey Department of the Research Center on the Rural Economy (RCRE) in

Beijing. We utilize household-level surveys from over 100 villages in 9 provinces (Anhui, Gansu, Guangdong, Henan, Hunan, Jiangsu, Jilin, Shanxi and Sichuan), supplemented with corresponding village-level surveys.[4] In each province, counties in the upper, middle and lower income terciles were selected, from which a representative village was then chosen. Depending on village size, between 40 and 120 households were randomly surveyed in each village. The survey spans the period 1986 to 1999, and includes between 7,000 and 8,000 households per year.

The RCRE originally intended a longitudinal survey, following the same households over time. While there is a household-level panel dimension to our sample, we observe considerable attrition of dhouseholds over the 1986-1999 period, especially after years when there was no survey. The RCRE was unable to conduct the survey in 1992 and 1994 because of funding difficulties. Households lost through attrition were replaced on the basis of random sampling.[5]

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

In Table 1, we report descriptive statistics on the composition of household income for 1987 and 1999. Over this period, per capita incomes rose from 578 RMB to 714 RMB, or an annual rate of slightly less than two percent. Growth up through 1996 was actually considerably higher, but household incomes declined beginning in 1996, largely as a result of a sharp fall in farm prices. These data capture the significant structural changes occurring in China's rural economy. At the start of the period, agriculture, which includes cropping income plus income from farm sidelines such as animal husbandry, was the most

---

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

[5] The issue of household attrition is examined in some detail in Benjamin, Brandt and Giles (2003).

important source of income for these households. Nearly every household reports positive farm income, and slightly more than half of household total net income is linked to farming. Wage income was the next important source, with more than a quarter of all income from wages, and more than two-thirds of all households reporting income from this source. This was followed by income from family run businesses (15.7%), and family transfers (5.0%). By the end of the period, on the other hand, the role of agriculture declined considerably, with less than a third of all income now coming from agriculture. Note also that the absolute income from agriculture declined. This sharp reduction was more than offset by rapid growth in family-run businesses and wage incomes, both of which increased at nearly five percent per annum over the entire period.

Table 2 provides summary level data on income and consumption inequality for our sample of households for four "benchmark" years, 1987, 1991, 1995 and 1999, along with the corresponding estimates of real per capita incomes and consumption. Over the entire period, the Gini coefficient for income inequality increased significantly from 0.32 to 0.37. In fact, the Gini was slightly pro-cyclical, increasing through the mid-1980s, falling with post-Tiananmen retrenchment for several years, increasing again between 1991 and 1996, and then falling the last three years for which we have data. As is usually the case, consumption inequality is lower, but it tracks the behavior of income inequality.

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

3.2 The Village-Level Data

We use the household-level data for the 12 years that the survey was carried out between 1986-1999 to construct a four period village-level panel. The periods are defined as follows: Period 1, 1986-1988, Period 2, 1989-1991, Period 3, 1993-1996, and Period 4, 1997-1999. For each of these periods, we calculate three-year averages of our variables. This method helps to smooth out some of the effect of transitory shocks in our income measures, and reduce other kinds of measurement error that may be present. All nominal variables are converted into "real" 1986 prices.

The existing literature examining links between growth, inequality and institutions has been largely conducted at the country level. Long-run analysis at a sub-country level, e.g. state or county, is predicated on heterogeneity at the sub-aggregate level in the initial level of inequality and institutions, and ultimately there being a significant potential role for local inequality in overall inequality.[6] In the analysis that follows, we take the "village" as our unit of observation. To help put this in perspective, a "typical" village in rural China consists of 300-400 households, and has a population of around 1500.

Our use of the village as unit of observation can be justified on several grounds. First, over the period we are analyzing, the village has played an important role as a distinct administrative entity.[7] Although the township (or former commune) is currently the lowest level of government in the state administrative hierarchy, major administrative functions are decentralized to the village, just as they were under the old collectives. Village government, for example, has responsibility for the allocation and management of key village resources, tax collection, and the provision of some public goods. Tax collection efforts, for example, include both the collection of the national land tax and farm output quotas, as well as the setting and collection of local taxes and fees. Constraints on household mobility imposed by the household registration system over a significant portion of this period may have strengthened the effect of village policy on household choices.

[6] See Panizza (2002) for an analysis of state-level inequality and growth within the U.S.
[7] Village boundaries have also been fairly fixed, and usually date back to the pre-1949 period

In rural China, land is not private-owned; rather, ownership resides at the collective or village level.[8] At the time of decollectivization, village leaders, usually in consultation with village households, played a pivotal role in deciding how to allocate usufruct rights to farmland among households in the village. They also played an important role in either the disposition or management and development of other collectively-owned assets, including village-owned enterprises, fish ponds, and forest land. Villages have also figured prominently, especially in the 1980s, in the provision of local public goods such as primary education, irrigation, drinking water, health facilities, and access roads.

Over the last twenty-five years, and despite some effort on the part of the central government to curb it, village governments have continued to exercise this authority over the allocation of key village resources, and have retained tax-setting powers. Their behavior can indirectly and directly influence the returns to household investment and labor supply. This suggests that local government policy may be an important factor in explaining growth at the village level, and also raises the possibility that inequality may be influencing growth through its effect on the local political economy. The likelihood of the latter has been enhanced by the introduction of elections in these villages in the late 1980s (Li and O'Brien, 2000), and the role of these elections in shaping village governance and policy. Previously, most of the authority resided with the village party secretary, who is appointed by the township level government. Interestingly, considerable heterogeneity appears to exist at the village-level in how far clean, competitive elections have advanced (Brandt and Turner, 2003). At the same time, however, the relaxation of restrictions on individual mobility, i.e. a weakening of the *hukou* system, market liberalizing reforms, and some re-centralization in the fiscal system may be having the opposite effects on the extent of the influence of villages on local growth prospects.

Second, over the period we are examining, and contrary with much of the conventional wisdom about China, a significant portion of the inequality in rural China is the product of differences in incomes between households within villages, or "within village inequality". Simple spatial decompositions that we

[8] In some cases, ownership resides at the sub-village level, namely, the small group, which in the pre-reform period corresponded to the old production team.

have carried out (see Benjamin, Brandt, and Giles, 2003) and also report in Table 2, show that in the mid-1980s, half of all of the inequality in rural China can be attributed to within-village income differences between households.[9] Three-quarters are the product of differences within provinces. The remainder is the product of differences in mean incomes across villages (or provinces). Moreover, by the end of the period, the contribution of within-village inequality to overall inequality had increased to nearly sixty percent, as the role of between-village differences in overall inequality actually dropped. The latter suggests that there may have been some convergence in incomes across villages over this period.

Within-village inequality is not only a significant portion of overall inequality, but there are also considerable differences across villages in the level of inter-household inequality. Figure 2 provides the histogram of the distribution of within-village income inequality for our sample of 104 villages for the beginning of our period 1986-1988 (top half) and the end of the period, 1997-1999 (bottom half). Initially, the mean (median) Gini coefficient for village income inequality was 0.23 (0.22), with considerable differences across our sample of villages. The 75^{th} (25^{th}) percentile was 0.27 (0.19). Over time, we observe that the entire distribution shifted right-wards, with both the mean and median village-level inequality rising to 0.27 and the dispersion across villages increasing only slightly (the standard deviation increased from 0.06 to 0.07).

A number of factors likely underlie these initial differences in inequality across villages including how resources such as land and human capital were distributed within villages. Although farm land was typically allocated on a very egalitarian basis in a village with the economic reforms in the late 1970s, early 1980s (Burgess, 1997, and Putterman 1993), differences in land endowments between production teams (small groups or *xiaozu*) within a village, and differences across villages in how they modified strict egalitarian rules for differences in the demographic composition of the household and off-farm labor supply, gave rise to differences in the overall distribution of this endowment among households. A typical village consists of 8-12 small groups. Land allocation decisions are decided at the village level,

---

[9] Similar decompositions we have done using other data corroborate the significant role of within village inequality in overall inequality.

but the actual distribution of land among households typically occurs at the level of the small group. In other words, the land of the small group is only allocated to households that make-up that group, and there does not appear to be much redistribution across small groups. Thus, differences in land endowments among these small groups can translate into differences across villages in land inequality.

Villages also differed in how they dispensed with land resources related to forestry, aquaculture, and animal husbandry, as well as collectively-owned physical assets, including small businesses that were previously run at the level of the production brigade (village). In addition, we observe significant differences across villages in the proportion of productive assets that were collectively owned, much of which is a legacy of the pre-reform period. On average, less than a quarter of all physical assets were collectively owned in period 1, but the difference in the proportion between villages in the 25th and 75th percentile was almost a third of total assets (0.421 versus 0.103). Not unexpectedly, the role of the collective in the local economy declined over time as reflected in the drop to 0.126 in the percentage of total assets that were collectively owned.

Finally, there are important differences across our sample of villages in the behavior of income and inequality over the entire period. As noted above, per capita incomes grew at slightly less than two percent per annum over the period. In a quarter of villages, however, growth was more than two times the sample average, while nearly a quarter of the villages experienced zero or negative annual growth between 1986-1999. These differences emerged, however, in the context of a general convergence in per capita incomes across villages. In Figure 1, we graph the average rate of growth of per capita incomes over the period against initial incomes, along with an ordinary least squares, and lowess regression through the points. The relationship is clearly negative.

Similar patterns are observed with respect to changes in the village-level Gini. There are considerable differences across villages in the change in inequality over the period, which increased on average .05 (from 0.22 to 0.27). These changes, however, are strongly negatively correlated with the Gini at the start of this period, as reflected by Figure 3. Note, however that the raw correlation at the village level between per capita income growth and the change in inequality over this period is zero (actually, 0.01, with p = 0.938).

4.0 Empirical Framework

4.1 Panel Data Specification

Our starting point is the dynamic panel data specification, employed (for example) by Forbes (2000):

$$\Delta \ln y_{it} = b_1 G_{it-1} + b_2 y_{it-1} + b'_3 X_{it-1} + h_t + a_i + u_{it}$$

where $\Delta \ln y_{it}$ is the growth in village i (per capita) income from period t-1 to t; G_{it-1} is the Gini coefficient in the initial period (t-1); y_{it-1} is the initial income level in period (t-1); X_{it-1} is a set of covariates (dated period (t-1)); h_t is a year effect common to all villages; a_i is a village fixed effect; and u_{it} is a well-behaved error term.

Ideally, we wish to estimate b_1, the "pure" effect of inequality on growth, controlling for all other factors that affect growth that may be correlated with inequality. For those components of X_{it-1} that we cannot measure, and which are constant over time, the inclusion of village fixed effects may help to net these out. This includes "fixed" institutional differences across villages. That said, a less-ambitious interpretation of b_1 may be more defensible. In this case, we wish to know if high inequality has any predictive power for subsequent growth, even if it is picking up the effects of unmeasured institutions or economic structure. The inclusion of lagged income helps to eliminate concerns that economic convergence will mechanically generate a negative relationship between inequality and growth. This could happen, for example, if poorer villages were more equal to begin with, but experienced more rapid growth. The convergence parameter (b_2) is of interest in itself, and also tells us something about the behavior of inter-village differences in income over time.

Besides omitted variables bias, there are several other econometric issues to consider. Foremost among these is measurement error, especially the possibly complicated relationships between noisy measures of village mean incomes, growth, and inequality. There are several possibilities.

- The coefficient on lagged income, b_2, will be biased downwards (i.e., more negative), showing a possibly greater level of convergence than truly exists;
- The Gini coefficient, and other measures of dispersion will be contaminated by the degree of measurement error, which may vary across villages. In any given year outliers in measured income may lead to exaggerated levels of inequality. Ignoring the possible correlation of this measurement error with growth rates, measurement error in the Gini will have the usual attenuation bias, making it more difficult to detect the genuine impact of inequality on growth;
- But of course, since inequality and growth are constructed with the same mis-measured household incomes, there may be more complicated biases than classical measurement error. For example, a single outlier may lead to a higher base level of income, and higher estimated inequality. In such a case, the lower growth associated with mean reversion of the measurement error would appear to be negatively correlated with the initial (over-stated) inequality, leading to a negative bias on the coefficient of inequality.

Indeed, one of our primary empirical concerns is to reduce the possibility that we are identifying a "mechanical" relationship between growth and inequality driven by measurement error. To address this, we employ two instrumental variables strategies that should at least break the possibly mechanical links between inequality and growth.

Our first strategy uses a semi-independent indicator of income levels and inequality, based on household consumption. While there is a common component to both household income and consumption (the value of home consumed agricultural output), we expect that consumption should be less subject to the kind of measurement error typical in household income data, especially concerning the measurement of "profits" in agriculture or family-run businesses. Furthermore, it is possible that consumption levels and inequality are more highly correlated with the long run distribution of income, dampening the short-run fluctuations of transitory income that may otherwise dominate the growth dynamics. We thus use lagged per capita consumption and the lagged Gini of per capita consumption as instruments for the corresponding income measures (G_{it-1}, y_{it-1}). Note that these instruments are motivated purely by measurement concerns, and will not address other forms of mis-specification like omitted variables bias, as we expect omitted institutional factors correlated with income inequality to be correlated with both income and consumption inequality.

The second strategy is even more removed from the measurement of income and inequality, and entails using measures of the level and distribution of village endowments, plus some indicators of village institutions. This set of instruments (Z_{it}) has the advantage over the consumption-based ones in that the measurement error in these variables is more likely independent of the measurement error in income. They also have the advantage of a possible "structural" interpretation, in that we may gain a better understanding of the underlying shifters of inequality that affect growth. These instruments may also help address simultaneity bias, whereby growth affects inequality. The down-side is that these instruments may be poorer predictors of the level of income and inequality than the consumption-based measures. An additional benefit of the instrumental variables specifications (using either set of instruments) is that we simultaneously address the dynamic panel data econometric issues, concerning the lagged endogenous regressor (as in Arellano and Bond).

4.2 Covariates and Instruments

Our covariates X_{it} are chosen with an eye to the macro growth/convergence literature, as well as the objective of controlling for institutional characteristics that may be correlated with inequality and growth. While there are many possible such variables available in the RCRE, we settle on a relatively sparse set of controls:

Education: This is measured as the log of the mean years of education per adult household member.[10] As in the growth literature, we expect it to be positively related to growth. Its inclusion may also clean up the convergence term (as in the convergence literature), as well as potential correlations between the level of human capital and the degree of inequality.

Proportion agriculture: This represents a very simple control for initial economic structure, and the extent of non-agricultural development, and is defined as the share of household income associated with crop production. It turns out that this is a potentially important control for exposure to agriculture price shocks, as crop prices fell significantly between periods 3 and 4. Agricultural villages were especially hard hit by crop price declines in the late 1990's, which led to both slower growth and increases in inequality. Inclusion of this variable helps partial the consequences of the price shock, and focus the regression more on the long run links between inequality and growth.

Market-quota price ratio: This is a potentially important institutional and/or policy variable, and is defined as the ratio of the market price of grain to the local administered, or quota, price of grain. The numerator will help detect differences in crop prices across villages (but mostly over time), while the denominator is correlated with the degree of taxation or distortion. The higher the ratio, the lower is the price received by farmers for the grain they are required to sell to the government (relative to the market price). We expect higher ratios to reflect higher degrees of distortion, and to possibly adversely affect growth.

In the panel specifications we also add village fixed effects, which will absorb the impact of other initial or relatively fixed institutional determinants of growth. In both the panel and cross-sections specifications we also experiment with provincial fixed effects, as well as unrestricted year effects.

As noted previously, we employ two classes of instruments: consumption-based, and endowment-based (Z). The consumption-based instruments are simply mean log per capita consumption, and the Gini

[10] More detailed descriptions of the variables are available in the (as yet unwritten) data appendix.

of per capita consumption. We partition the endowment-based instruments into "lean" and "extended" sets, in order to assess the robustness of results to a stronger set of exclusion restrictions:

a) Lean set of instruments

This set of instruments is based on the level and distribution of two factors of production: land and labor. The four variables are:

Mean cultivated land per capita: This instrument should help predict the initial level of income, and possibly be correlated with the level of income in each period.

Gini of per capita cultivated land per capita: Inequality of cultivated land should generate greater inequality of income.

Mean share of dependents in household: Dependents are household members who are outside of working ages or unable to work. Women over 55, men over 60 and children who are still in school or under age 16 are considered dependents. A higher fraction of dependents should be negatively related to the level of income per capita (as a greater share of the "per capita" is non-productive).

The Coefficient of variation of dependents per household: inequality of productive household members per capita should help predict the level of inequality.

Of course, we are also assuming that controlling for overall inequality and the initial level of income, these variables do not belong in the growth equation. While they have limits, we report the results of over-identification tests as a specification check.

b) Extended set of instruments

The extended instrument set includes the lean instruments, plus the following "capital" related variables:

Mean assets per capita: This variable is constructed from the village survey, and equals the real value of productive assets (agricultural and non-agricultural) in the village, normalized by population. This variable should be correlated with initial incomes.

Gini of per capita productive assets: Constructed from the household survey, this variable captures the inequality of accumulated productive assets per household. This variable should be correlated with income inequality.

Percentage of Assets Collectively Owned: This variable represents the fraction of village productive assets that is owned collectively by the village. This could reflect the degree of "collectivization" in the village, and ay be correlated with both inequality and initial incomes. It might also reflect village preferences for investment in public goods.

4.3 Cross-Section Specification

As argued by previous authors (e.g., Easterly, Barro), the panel-data specification, especially using fixed effects, identifies linkages between inequality and growth of short-run, high frequency covariation, which may be dominated by measurement error. Even with careful instrumenting, it may also be the case that the processes that we are attempting to identify only matter in the medium to "long run", in which case they will not be detectable in the panel. We thus also estimate a cross-section version of the model:

$$\Delta \ln y_i = b_0 + b_1 G_{io} + b_2 \ln y_{io} + b_3' X_{io} + u_i$$

where $\Delta \ln y_i$ is growth over the full period, and G_{io}, y_{io}, X_{io} are the initial village Gini, income level, and characteristics. This is basically the model used in the macro-growth literature (e.g., Persson and Tabellini) We estimate this model, employing the same strategies as with the panel data. The only difference is that the instruments and covariates are defined according to the initial period, as opposed to "t-1".

4.4 Reverse Specification

Our focus is on the linkages between inequality and growth, where causality runs from inequality to growth. However, previous researchers have also highlighted causality going the other direction, from growth to inequality. For example, the structural changes associated with development (growth) may also lead to higher inequality, at least in the short run (e.g., Kuznets (1955); Barro (2000)). We make no attempt to identify this relationship, though we do highlight the results of estimating the regression this way, especially in underlining the potentially positive covariation of changes in inequality and growth that may confound the causal link from inequality to growth. The panel and cross-section specifications are given by:

$$\Delta G_{it} = a_0 + a_1 \Delta \ln y_{it} + a_2 G_{it-1} + a_3' X_{it-1} + e_{it}$$

$$\Delta G_i = a_0 + a_1 \Delta \ln y_i + a_2 G_{i0} + a_3' X_{i0} + e_i$$

In this model we maintain the same covariates – acknowledging that a more complete model would have regressors specific to a model for the evolution of inequality. We employ similar estimation methods for the reverse specification as the main one, except that we spend less time worrying about the endogeneity of $\Delta \ln y_i$ -- not because it is unlikely to be an issue – but because we are not interested in the structural (causal) interpretation as much as the empirical covariation. We do however instrument the initial inequality level with our usual set of instruments, and we also use the consumption-based measures as instruments for $\Delta \ln y_i$ in order to assess the possible consequences of measurement error in dampening (or amplifying) the Kuznets relationship.

5.0 Results

5.1 Panel Specifications

a) Without instrumenting

In Table 4 we begin with the panel data specification, exploring the sensitivity of the "OLS" specification to the inclusion of covariates. All the results in this table employ the random effects estimator. The first four columns estimate the model without "X's", variously adding period and province fixed effects. Our focus is the coefficient on the lagged Gini. Note first however, that the results show a significant negative correlation between growth and initial income, which is consistent with income convergence across villages, introduced previously. As in the more general growth literature, the degree of convergence is greater when we include more controls, especially province effects. As for the Gini coefficient, it is generally small and insignificant, and usually positive. In columns (5) through (8), we add the X's, and repeat the exercise of variously adding the province and period effects. The basic story is not sensitive to province or period effects, and as our preference is to err on the side of "over-controlling" (when possible), we will treat the specification with both province and period effects as our base. In column (8), we thus see the familiar pattern of income convergence (higher with, than without controls), and no evident link between lagged inequality and growth. Concerning the controls themselves, we see that while the fraction of income from agriculture is negatively related to future growth, this effect becomes insignificant with the inclusion of province and period effects. The market-quota price ratio is also insignificant, essentially in all specifications. The most robust and significant control variable is education, which is positively related to growth. The results in table 4 thus bear striking resemblance to the cross-country literature, showing convergence (increasing once we control for education), a strong positive link between education and growth, and none between inequality and growth.

b) Employing instrumental variables

In Table 5, we introduce two innovations to the specification, incorporating fixed effects, and employing instrumental variables. All specifications include the covariates from column (8) of Table 4 (with province and period effects). Columns (1) and (2) compare the non-instrumented random versus fixed effects estimates (Column (1) is identical to Column (8) from Table (4)). Clearly, fixed effects make a huge difference for the convergence parameter, though one expects this is driven largely by measurement error. The coefficient on inequality becomes more positive, though it remains insignificant. This increase in the coefficient is not what one would expect with simple classical measurement error in levels of the Gini. However, one possibility is that the short-run covariation of inequality and growth is dominated by the positive "Kuznets" relationship, and this covariation is amplified in the fixed effects specification. The Hausman test of FE versus RE strongly rejects that the two estimators estimate the same parameter vector, though this rejection arises primarily in the difference between the lagged income coefficient, which is most dubious in the fixed effects specification (for both measurement error, and dynamic panel data reasons).

Columns (3) and (4) report results using the consumption-based instruments. In the Random Effects specification, the lagged income coefficient is a little smaller than without IV, as one would expect if the convergence coefficient was negatively biased by measurement error. The inequality coefficient, however, is virtually identical and insignificant. The Hausman test of IV versus no IV leads us to conclude that the no-IV specification should be rejected, though this rejection is not driven by differences in our key coefficients. Throughout the table, we also report the F-statistic for the joint significance of the identifying instruments in the first-stage regressions, in order to minimize concerns over weak instruments. The instrumented FE specification is quite different from the non-instrumented version, at least at first look. As with RE, the convergence parameter is lower when instrumented, though it is negative and significant. Most surprisingly, the instrumented Gini coefficient is now *positive* and significant, and significantly different than the non-IV estimate. If true, this suggests that inequality is positively related to growth, and that increases in inequality are followed by period of higher growth. While we cannot entirely dismiss this coefficient, it turns out to be an aberration compared to all our other results. While this requires more investigation, our preliminary understanding is that the fixed effects specification is especially sensitive to big changes in consumption inequality, and inclusion of these observations seem to be responsible for the large positive coefficient. Certainly, this result is not robust to variations in specification, and the Hausman test also suggests that the IV specification is (jointly) no different than the non-IV version.

In the remaining four columns we use the economic characteristics of the villages as instruments, beginning with the "lean" set (land and labor). The random effects estimates are similar to both the consumption-based IV, and "OLS" estimates, showing significant convergence, but no relationship between inequality and growth. In these specifications, while insignificant, the estimated coefficients tend to be negative. Most interestingly, the basic results in these (4) columns suggest that neither fixed effects or instrumental variables make much difference in the basic conclusions, over the standard non-instrumented equation in column (1). It is also important to note that the instrumental variables estimators themselves satisfy the usual battery of specification tests, having significant identifying instruments in the first stage regression, and passing the overidentification tests.

c) Testing for non-linearities

In Table 6 we explore whether there may be non-linearities in the relationship between the lagged Gini, and subsequent growth. This partially addresses issues raised by Banerjee and Duflo (2003), who show that non-linearities exist in the relationship between changes in inequality, and growth. They are able to capture non-linearities with a quadratic term. We also add a quadratic term for the Gini, though our specification is in levels, not changes. Our closest specifications to theirs will be those with fixed effects, which come closest to first differences. In addition to the inclusion of the quadratic terms, we instrument lagged income and both quadratic terms. For the consumption-based instruments we add the squared consumption Gini to the set, while the extended set remains the same. In none of these specifications do we find a significant link between inequality and growth, either for individual coefficients, or jointly for both terms of the quadratic. At least within this specification, it does not appear that our "non-result" is the consequence of over-linearizing a non-linear relationship.

d) Estimating the "Kuznets" specification

One explanation of the zero-coefficient on the Gini could be simultaneity bias. If a Kuznets' equation belongs in our system, with inequality increasing as a consequence of economic development, then our estimator will be positively biased, as noted by Lundberg and Squire (2003) for example. One potential way to address this is through instrumental variables, not using the consumption-based instruments, which are driven by the same processes, but by the endowment instruments. As long as we maintain that those instruments can be excluded from the growth equation, then we can identify the impact of inequality on growth, even with simultaneity bias. As previously noted, such instrumenting makes no difference.

In Table 7 we report results of estimating the equation in reverse, to evaluate how strong a positive "Kuznets" relationship may exist in the data, possibly confounding our effect of inequality on growth. In this specification, we regress the change in inequality on initial inequality and contemporaneous growth. We have a limited set of instruments with which to identify the "pure" effect of growth on inequality, and it is not the main objective of our paper. However, we do employ our existing instruments which allow us to explore possible consequences of measurement error. Looking across the columns of Table 7, two main results appear. First, there appears to be strong convergence in the level of inequality across villages. Second, there is a robust "non-relationship" between changes in inequality and contemporaneous growth. Given the nature of the instruments, we do not believe this is driven by measurement error. Quite simply, there is no detectable relationship between inequality and growth, with causality in either direction.

5.2 Cross-section Specifications

a) Main cross-section results

In Table 8 we move to the cross-section specification, which is a potentially better arena in which to detect the long run, or slowly evolving, relationship between inequality and growth. The dependent variable is average growth between periods 1 and 4, while the regressors are all values of the variables in period 1. Before discussing the regression results, in Figure 4 we show the bivariate relationship between initial inequality and subsequent growth. The relationship is basically negative, but statistically insignificant. The non-parametric regression is also essentially linear through most of the observations, and it does not appear that the overall negative result is driven by the two or three "outlier" villages with initial Gini's around 0.40. In Table 8, the specifications vary in terms of being OLS or instrumented (using either consumption or the extended set of instruments), and with or without province effects.

In column (1), we show results of the base regression with no covariates or province effects. As in the panel, we find strong evidence of income convergence, and no evidence linking inequality to growth (as suggested by Figure 4). However, in columns (2) and (3), once we add controls and province dummies, we find a significant negative coefficient on initial inequality. Furthermore, the additional regressors are themselves significant. First, the share of income from agriculture is significant and negative. This does not arise from the more agricultural villages being poorer, as we are controlling for initial incomes. Instead, it suggests that those villages more heavily engaged in agriculture at the beginning, ended up growing more slowly. To some extent, we will see that this is due to greater exposure to the collapse of crop prices in period 4. Second, the market-quota price ratio is negative, suggesting that in villages where farmers faced more "distorted" crop prices, growth was slower. We cannot tell whether this reflects broader distortionary consequences of local government behavior. Besides, its effect basically disappears once we add province dummies. And third, initial education is significantly positively related to growth.

In columns (4) and (5) we address the possibility that measurement error and/or some other mechanical link between inequality and growth is driving the OLS conclusions, by using the consumption-based instruments. The results are basically the same, yielding essentially larger estimates of the adverse impact of inequality on growth, consistent with measurement error and attenuation bias in the OLS specification. However, the Hausman Test fails to reject the null hypothesis that the OLS equation is properly specified. In Columns (6) and (7) we report the corresponding results using the extended instrument set. This specification yields the largest negative coefficients for initial inequality, though the Hausman tests indicates that (as a vector) we find no evidence to prefer the IV results.

It is instructive to explore the IV specification in Column (7) more closely, as the identifying assumptions yield some potentially important insights into the process linking inequality and growth. In Table 9 we report the reduced form equations for each endogenous variable (initial inequality, initial income, and the dependent variable, growth). All specifications here include province effects. Turning first to column 2, we see that the levels of assets per capita and the dependency ratio are good predictors of initial income (in the expected way). The excluded instruments are also jointly significant. In column (4), the first stage regression for initial inequality shows only one significant excluded instrument, the percentage of assets that is collectively owned being negatively related to inequality. Interpretation of this coefficient is not obvious. However, one possibility is that villages with higher rates of collective ownership have other egalitarian-oriented institutions. Or, possibly these villages have a greater taste for investment in public goods. Whatever the interpretation, it appears that these villages also have lower levels of initial inequality. In column (6), we also confirm that such villages experience higher growth, controlling for a variety of other characteristics. The IV estimate from column (7) of Table 8 is essentially identified off this relationship. While this specification passes the over-identification test, and the excluded instruments do not appear (as a group) to be weak, there always lingers the possibility that this IV specification is also biased because of some unobserved heterogeneity. However, one remarkable feature of Table 8 is the robustness of the negative relationship between growth and inequality whatever the instrument set chosen. That such different exogeneity assumptions should yield similar answers informally suggests that the relationship passes a "grander" overidentification test.

b) reconciling the panel and cross-section results

So we have two competing sets of results. From the panel we find no relationship between inequality and growth, but from the cross-section we find a robust negative one. Possibly, the high-frequency variation is the source of the discrepancy. One simple way to investigate this is to see to what extent we can mimic the cross-section results within the panel data framework. We take two approaches. First, we use initial values of the instruments (not period specific) in instrumental variables estimation. This is essentially the strategy adopted by Lundberg and Squire (2003), though they cast their results in a dynamic panel data setting. The second approach is to present the "Between Estimator", which regresses long run levels of variables (like average growth rates) on village mean characteristics. The results of these exercises are presented in Table 10. In columns (1) and (2) we show the non-instrumented results. While Between estimates are negative, these results look a lot like the previous panel data results. Using the initial consumption-based instruments, we estimate results that look more like the cross-section, though initial inequality is still insignificant. However, using the extended set of instruments, we estimate statistically significant negative coefficients for initial inequality. Thus, even within the panel data framework, if we use non-time varying instruments (initial conditions for the villages), we obtain similar results to the cross-section. Such instruments will accentuate long run correlations in the data, and it appears that one needs to "tune into" that lower frequency in order to detect the relationship.

c) The impact of inequality on sectoral development

If we accept that higher initial inequality is correlated with lower subsequent growth, and that this is not an artifact of measurement error, the obvious question is why? Which (if any) of the potential theories linking distribution to growth underlies the relationship in Chinese villages? In principle, we could imagine that the key unobserved variable (explanation) is given by q_{it}, which captures the access to credit facilitated by greater equality, or the provision of better public goods like quality education (to choose two candidate explanations). Ideally, with data or proxies for q_{it}, we could add them to the model and have the inequality effect disappear, thus "explaining it away." Unfortunately, even with the rich set of covariates we have, that is probably an unattainable standard of evidence, given the inherently "difficult to observe" nature of the candidate q_{it}'s. That said, by observing more correlations between inequality and other institutional or economic outcomes, we may be able to narrow down the types of q_{it} relevant for the Chinese context.

As a starting point we explore linkages between inequality and other indicators of "economic development" besides average household income. Specifically, in Table 11 we disaggregate household income into sub-components in order to see which parts of income were most adversely affected by initial inequality. The regressions are slight variants of the cross-section specifications in Table 8:

$$\Delta \ln y_i^{Type} = b_0^{Type} + b_1^{Type} G_{io} + b_2^{Type} \ln y_{io}^{Type} + b_3^{Type\prime} X_{io} + u_i^{Type}$$

The primary difference is that we control for initial income of the particular type, though inequality is still measured as the Gini for overall income inequality. We estimate the regressions by OLS and Instrumental Variables, employing the extended set of instruments, and report specifications with province fixed effects.

In the first two pairs of columns we break income into "Agricultural" and "Non-agricultural". One possible avenue by which inequality could affect economic growth is through credit market imperfections. For example, if retained earnings are the only way to finance investment in family non-agricultural businesses, and assuming that there are minimum thresholds for such investments, then for a given level of average income we would expect higher possible levels of investment in more equal villages. Columns (1) and (2) show OLS and IV estimates for agricultural income. In the OLS specification, inequality is *positively* related to agricultural growth. Basically, villages with higher levels of inequality in period one appear to have experienced greater specialization in farming. This is especially apparent by comparing the agriculture results to non-agricultural income growth, which is significantly negatively related to initial inequality. Also note that education plays an important role in tilting growth towards non-agricultural sources. While positive and insignificant for agriculture, it is larger and significant for non-agriculture. These results are consistent with the credit-market explanation, but also with the tendency of more equal villages to invest in public infrastructure that fosters non-agricultural income growth, especially education.

One way to distinguish between these explanations is to further disaggregate non-agricultural income in order to highlight the role of inequality in the development of family businesses. In the next two pairs of columns we show results for wage income (including household members working outside the village) and family-run businesses. Most surprising (to us) almost all of the effect of inequality seems to operate through wage income, not family businesses. Given the important role of education in accessing off-farm jobs (as is clear in columns 6 and 7), this result is suggestive that the "public good" or an education-based explanation may link inequality to growth. This may also be supported (with a stretch) by our previous finding linking greater collective ownership with lower inequality and higher growth. It would be difficult to draw a direct link between greater collective ownership of assets and access to non-agricultural employment, and instead it may proxy for political economy mechanisms. Note that this result does not preclude the possibility that credit is part of the story: it may be necessary for households to finance off-farm employment (including migration) through "retained earnings." However, if that was a key consideration, we would expect to see inequality mattering more in the family-run business equation.

d) The cross-section Kuznets regression

Returning to results concerning total income growth, the remaining two tables explore other issues of specification and robustness. In Table 12 we estimate the cross-section variant of the “Kuznets” relationship, in order to see whether the long-run perspective yields a stronger link between growth and changes in inequality. We estimate a number of specifications, with and without province effects, by OLS and instrumental variables. In addition, we estimate a median regression in order to see whether our result is driven by outliers. In short, we find no significant evidence of a Kuznets relationship, whereby faster growing villages experienced larger increases in inequality. First, this suggests that the positive Kuznets relationship may not be important in the Chinese context, so that estimation of the causal link between inequality and growth is less contaminated by simultaneity. Second, it also suggests that the positive co-movement between inequality and growth at the aggregate level is not causally linked at the village level: the fastest growing villages are not the ones experiencing the greatest increases in inequality.

e) Robustness to time period

Finally, in Table 13 we investigate the sensitivity of our conclusions to the definition of “long run.” As emphasized in our previous paper (Benjamin, Brandt, and Giles (2003)), the fourth period is characterized by significantly widening income disparities (within villages), as well as low growth, driven mostly by collapsing grain prices. Thus, between periods three and four, agricultural villages experienced increases in inequality combined with lower growth. This is the opposite direction of the Kuznets relationship, and may explain why do not find it in Table 12. In Table 13 we estimate our key regressions (by OLS only) with the endpoint of the growth period taken as period 3, essentially restricting our analysis to the high growth periods. In column (1) we report the basic cross-section growth results, and see that initial inequality still has a significant negative impact on growth, though the magnitude of the coefficient is somewhat smaller than the 4-period growth results. This suggests that while growth was high, the adverse impact of inequality on growth was lower. Columns (2) and (3) show why. Here, the impact of initial inequality on the agriculture/non-agriculture split is even more pronounced. For whatever reason(s), high inequality villages significantly tilted towards agricultural growth, at the expense of non-agricultural development. While crop prices were high, this had little consequence on overall growth. However, with the collapse of crop prices over the late 1990's, these exposed villages simultaneously suffered drops in average incomes, and experienced widening income inequality as farm households became relatively poor (within these villages). Finally, in column (4), we report the results of the "Kuznets" specification, where we now see a significant positive relationship between inequality and growth. It thus appears that this positive relationship exists in the high-growth period, and is obscured when we include the period three to four episode. During the high growth period, faster growing villages were becoming more unequal.

6.0 Conclusions

Distilling a large household-level data set spanning the post-reform period from 1986-1999 into a panel of approximately 100 villages covering the same period, we estimate whether inequality affects growth. Using cross-period covariation of inequality and growth (i.e., the standard panel data specification), we find no relationship between inequality and growth. Our conclusions are robust to a number of instrumental variables strategies, so we are confident that our results show there is no reliable link between inequality and growth at high frequencies. However, when we switch to a cross-section, "long-run," summary of the growth experience of the villages, we find robust, statistically significant evidence that inequality reduces growth. Taking -0.10 as a representative magnitude of the effect, this implies that a difference in village Gini of 0.10, e.g., from 0.15 to 0.25, results in average annual growth rates lower by one percent per year, or half of the average growth rate over this period.

While it is asking too much of the data to pin down the actual mechanism by which inequality exerts its negative effect, we do find that inequality strongly tilted village economic activity away from higher growth non-agricultural development, and towards agriculture. It may be that higher inequality proxies for poor access to capital, or poorer local public investment in infrastructure (like education) that facilitates the movement of households out of agriculture. In light of the recent steep increases in inequality experienced by the most agricultural villages, our results raise important policy questions concerning the longer run impact of current inequality on future rural economic growth prospects.

7.0 References

Alesina, Alberto and Dani Rodrik. 1994. "Distributive Politics and Economic Growth." *Quarterly Journal of Economics*, 109:2, pp. 465-90.

Anand, S. and Ravi Kanbur. 1993. "The Kuznets Process and the Inequality-Development Relationship." *Journal of Development Economics*, 40, pp. 25-52.

Arellano, Manuel and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58:2, pp. 277-97.

Banerjee, Abhijit and Esther Duflo. 2003. "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth*, 8, pp. 267-99.

Barro, Robert J. 2000. "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth*, 5, pp. 5-32.

Benabou, Roland. 1996. "Inequality and Growth," in *NBER Macroeconomics Annual*. Ben S. Bernanke and Julio Rotemberg eds. Cambridge, MA: MIT Press, pp. 11-74.

Benjamin, Dwayne and Loren Brandt. 1997. "Land, Factor Markets, and Inequality in Rural China: Historical Evidence." *Explorations in Economic History*, 34:4, pp. 460-94.

Benjamin, Dwayne, Loren Brandt, and John Giles. 2003. "The Evolution of Income Inequality in Rural China." Working Paper, William Davidson Institute.

Brandt, Loren and Matthew Turner. 2003. "The Usefulness of Corruptible Elections." Manuscript, University of Toronto.

Burgess, Robin. 1997. "Land, Welfare and Efficiency in Rural China." Manuscript, London School of Economics.

Deininger, Klaus and Lyn Squire. 1996. "A New Data Set Measuring Income Inequality." *World Bank Economic Review*, 10:3, pp. 565-91.

Easterly, William. 2002. "Inequality does cause underdevelopment: New evidence." Center for Global Development, Working Paper #1.

Engerman, Stanley and Kenneth Sokoloff. 2002. "Factor Endowments, inequality, and paths of development among New World economies." NBER Working Paper, #9259.

Forbes, Kristin. 2000. "A reassessment of the relationship between inequality and growth." *American Economic Review*, 90:4, pp. 869-87.

Kaldor, Nicholas. 1956. "Alternative Theories of Distribution." *Review of Economic Studies*, 23:2, pp. 83-100.

Kaldor, Nicholas. 1957. "A Model of Economic Growth." *Economic Journal*, 67, pp. 591-624.

Kuznets, Simon. 1955. "Economic growth and income inequality." *American Economic Review*, 45:1, pp. 1-28.

Lewis, Arthur. 1954. "Economic Development with Unlimited Supplies of Labor." *The Manchester School*, 22, pp. 139-91.

Li, Hongyi and Hengfu Zou. 1998. "Income Inequality is not Harmful for Growth: Theory and Evidence." *Review of Development Economics*, 48, pp. 600-21.

Li, Lianjiang and Kevin O'Brien. 2000. "Acommodating Democracy," in *Elections and Democracy in China*. Larry Diamond and Ramon Myers eds. New York: Oxford University Press.

Lundberg, Mattias and Lyn Squire. 2003. "The simultaneous evolution of growth and inequality." *Economic Journal*, 113, pp. 326-44.

Panizza, Ugo. 2002. "Income Inequality and Economic Growth: Evidence from American Data." *Journal of Economic Growth*, 7, pp. 25-41.

Perotti, Roberto. 1996. "Growth, Income Distribution, and Democracy: What the Data Say." *Journal of Economic Growth*, 1996:1, pp. 149-87.

Persson, Torsten and Guido Tabellini. 1994. "Is Inequality Harmful for Growth?" *American Economic Review*, 84:3, pp. 600-21.

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

Wang, Shaoguang. 1999. *The political economy of uneven development : the case of China*: M.E. Sharpe.

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

Table 1: Household-level Descriptive statistics
The composition of income in 1987 compared to 1999
(in 1986 RMB)

	1987			1999			Growth
	Mean	Share	% > 0	Mean	Share	% > 0	
Total Income	578	1.00	1.00	714	1.00	1.00	0.018
Agricultural Income	229	0.397	0.98	158	0.222	0.94	-0.031
Grain Income	175	0.303	0.98	113	0.158	0.93	-0.036
Cash Crop Income	46	0.080	0.81	30	0.042	0.56	-0.036
Fruits, Tea and Dates	8	0.014	0.25	15	0.022	0.26	0.052
Agricultural Sidelines	74	0.129	0.96	68	0.095	0.76	-0.007
Forest Products	17	0.029	0.33	10	0.014	0.19	-0.044
Livestock	54	0.093	0.95	50	0.069	0.75	-0.006
Aquaculture	4	0.007	0.12	8	0.012	0.06	0.058
Family Businesses	91	0.157	0.62	162	0.227	0.50	0.048
HH Industry	27	0.048	0.14	44	0.061	0.07	0.041
Construction	6	0.010	0.07	11	0.016	0.05	0.051
Transportation	17	0.030	0.06	26	0.037	0.08	0.035
Commerce, Service & Trade	25	0.042	0.13	57	0.079	0.17	0.069
Other Family Business Income	16	0.027	0.39	24	0.034	0.25	0.034
Wage Income	145	0.251	0.71	276	0.387	0.68	0.054
Local Wage Income	85	0.147	0.45	79	0.111	0.26	-0.006
Employment Outside	46	0.080	0.39	175	0.245	0.50	0.111
Local Gov't Employment	14	0.024	0.07	22	0.031	0.05	0.038
Family Transfers	29	0.050	0.52	34	0.048	0.49	0.013
Government Transfers	4	0.008	0.65	6	0.008	0.71	0.034
Other Income	5	0.009	0.14	8	0.012	0.11	0.039
SAMPLE SIZE	7,983			6,987			

Notes:
1) This table compares the composition of income in 1987 to 1999.
2) Average real per capita income is shown for detailed sub-categories of income, along with the share of total income ("Share") accounted for by that income, and the proportion of households with non-zero income for that source.
3) The last column reports the implied annual growth rate of income for that source.
4) Wage income is divided between "Local" wage income, and "Employment Outside". Employment outside is employment outside of the village, and in most cases outside the township. This category includes both family members employed in distant migrant labor markets and those who are commuters returning to the village on weekends.
5) Source: Authors calculations, reported in Table 5 of Benjamin, Brandt, and Giles (2003).

Table 2
Summary of key Household-level Outcomes Concerning Growth and Inequality
RCRE, Selected Years

	1987	1991	1995	1999
Real Per Capita Income	578	551	772	714
Real Per Capita Consumption	410	402	548	508
Gini Per Capita Income	0.32	0.33	0.33	0.37
Gini Per Capita Consumption	0.25	0.27	0.27	0.31
Sample Size	7,983	7,903	6,738	6,987
Spatial Dimensions of Inequality (Proportion of variation of log per capita income explained)				
Contribution of region	0.186	0.162	0.154	0.120
Contribution of province	0.237	0.218	0.183	0.153
Contribution of village	0.500	0.466	0.413	0.424

Notes:

1) This table shows mean household income and consumption for selected years, as well as the corresponding Gini coefficients.
2) This table also shows the fraction of variation of real log per capita income (and consumption) attributed to location. This is simply the R-squared from a regression of log per capita income on a set of location dummies.
3) The effect of location is reported at three levels of aggregation: (1) Region, defined as West (Gansu, Shanxi, and Sichuan), Central (Anhui, Henan, and Hunan), and East (Jilin, Jiangsu, and Guangdong); Province (nine provinces previously listed); and Village (number by year, for: 1987: 111; 1991: 113; 1995: 106; 1999: 112).
4) Source: Authors calculations, reported in Benjamin, Brandt, and Giles (2003), Tables 1,2,4.

Table 3
Summary of Village Characteristics: Beginning and Final Periods

	Period 1 (1986-87-88)			Period 4 (1997-98-99)		
	25th	Median	75th	25th	Median	75th
Key variables						
Per capita household income (y)	410	553	709	522	622	867
Period 1 to 4 growth rate of income				-0.32	1.87	3.66
Gini (income)	0.190	0.220	0.270	0.223	0.268	0.307
Change in Gini (period 1 to 4)				-0.007	0.041	0.076
Primary Covariates						
Education (Years per adult)	4.65	5.76	6.37	5.95	6.69	7.38
Crop income share	0.332	0.490	0.598	0.158	0.288	0.460
Market/Quota price ratio	1.20	1.40	1.56	0.91	0.98	1.08
Potential Instruments						
Per capita consumption	294	389	480	370	485	609
Period 1 to 4 growth rate of consumption				0.18	2.34	3.70
Gini (consumption)	0.148	0.168	0.186	0.177	0.203	0.240
Change in Gini (period 1 to 4)				0.013	0.033	0.062
Per Capita Cultivated land	0.88	1.22	2.17	0.75	1.06	1.69
Gini (cultivated land)	0.135	0.168	0.211	0.181	0.215	0.287
Per Capita productive assets	164	268	505	529	1017	2280
Gini (Per capita productive assets)	0.387	0.516	0.653	0.421	0.499	0.670
Share of dependents in household	0.405	0.438	0.476	0.342	0.379	0.433
CV of Dependents' share	0.357	0.431	0.491	0.464	0.565	0.671
Proportion of collectively owned assets	0.103	0.244	0.421	0.022	0.126	0.281

Notes:

1. This table reports descriptive statistics for the key village-level variables used in this paper, over the sample of 101 villages.
2. For each variable we report the median within the sample, as well as the 25th and 75th percentiles, in order to provide some indication of the variation across villages.
3. The "Key variables" are the growth and inequality variables that we focus throughout the paper. Primary covariates are the regressors (X's) included in virtually all specifications, while the potential instruments are those variables used as instruments in our IV specifications.

Table 4: Exploring the Panel Data Specification – Sensitivity to Covariates
Dependent variable: Village growth rate ($\Delta \ln y_{it}$)
Random Effects Estimator
(standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lagged Income (y_{t-1})	-0.128* (0.030)	-0.124* (0.029)	-0.192 (0.040)	-0.170* (0.037)	-0.180* (0.033)	-0.152* (0.030)	-0.254* (0.042)	-0.208* (0.038)
Lagged Gini (G_{t-1})	-0.015 (0.196)	0.066 (0.183)	0.133 (0.231)	0.197 (0.208)	-0.152 (0.194)	-0.060 (0.174)	0.046 (0.232)	0.130 (0.205)
Lagged Crop Share					-0.188* (0.065)	-0.161* (0.058)	-0.121 (0.072)	-0.098 (0.064)
Lagged Market-Quota Price Ratio					-0.036 (0.025)	-0.041 (0.023)	-0.025 (0.028)	-0.026 (0.024)
Lagged Log Household Education					0.142* (0.056)	0.125* (0.051)	0.242* (0.071)	0.213* (0.064)
Year Effects?	No	Yes	No	Yes	No	Yes	No	Yes
Province Effects?	No	No	Yes	Yes	No	No	Yes	Yes

Notes:
1. Estimated over the sample of 101 villages, for 4 periods (yielding three observations per village).
2. All specifications include a constant. Year dummies and/or province dummies also included where indicated.
3. "*" indicates statistically significant at the 5% level.

Table 5: Instrumental Variables Estimates of Panel Data Specification
Varying the instrument set, Fixed and Random Effects
Dependent variable: Village growth rate
All Specifications with Year and Province Effects, Standard Covariates
(standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	No Instruments		Instrument Set: Consumption		Instrument Set: "Lean"		Instrument Set: "Extended"	
	RE	FE	RE	FE	RE	FE	RE	FE
Lagged Income (y_{t-1})	-0.208* (0.038)	-0.791* (0.075)	-0.143* (0.049)	-0.444* (0.178)	-0.197* (0.080)	-0.602 (0.385)	-0.160* (0.064)	-0.528 (0.321)
Lagged Gini (G_{t-1})	0.130 (0.205)	0.563 (0.310)	0.113 (0.338)	1.748* (0.786)	-0.476 (0.677)	-0.416 (1.065)	-0.494 (0.451)	-0.052 (0.739)
Hausman Test: RE vs. FE	85.29* (0.0000)		8.86 (0.263)		4.36 (0.738)		4.03 (0.78)	
Hausman Test: IV vs. No IV			142.11* (0.0000)	7.97 (0.335)	-2.53 (N/A)	1.14 (0.992)	-0.54 (N/A)	1.08 (0.99)
F-test for excl. instruments (y_{t-1})			201.50* (0.0000)	27.89* (0.0000)	23.23* (0.0000)	2.04 (0.09)	27.14 (0.0000)	2.02* (0.05)
F-test for excl. instruments (G_{t-1})			78.81* (0.0000)	22.12* (0.0000)	7.58* (0.0000)	4.79* (0.0011)	12.32* (0.0000)	7.78* (0.0000)
Overid test (p -value)					2.389 (0.303)	0.736 (0.692)	4.231 (0.517)	5.941 (0.213)

Notes:
1. All specifications include the standard covariates (lagged crop share, market-quota price ratio, and log household education), plus year and province effects.
2. The IV specifications vary by instrument set. In each IV specification, the endogenous regressors are lagged income and the lagged Gini. Columns (3) and (4) use the level and Gini of per capita consumption as identifying instruments; Columns (5) and (6) use the "lean" instrument set: per capita cultivated land, the Gini of per capita land, and the mean and CV of the dependency share; Columns (7) and (8) extend this set by adding the mean and Gini of per capita assets, plus the percentage of village assets that is collectively owned.
3. The Hausman tests (and p-values) are reported for a variety of comparisons as indicated (i.e., fixed versus random effects; IV versus no-IV).
4. The F-tests for excluded instruments test the joint significance of the excluded instruments in the first stage regressions.
5. The Overidentification test (and p-value) is also reported for the over-identified specifications.

Table 6: Is there a quadratic effect of the Gini?
Panel Data Specifications, Fixed and Random Effects
Dependent variable: Village growth rate
All Specifications with Period and Province Effects, Standard Covariates
(standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
	No Instruments		Instrument Set: Consumption		Instrument Set: "Extended"	
	RE	FE	RE	FE	RE	FE
Lagged Income (y_{t-1})	-0.205*	-0.782*	-0.141*	-0.440*	-0.165*	-0.709
	(0.038)	(0.077)	(0.049)	(0.177)	(0.066)	(0.448)
Lagged Gini (G_{t-1})	1.396	1.223	2.664	-0.164	0.884	-3.23
	(1.023)	(1.274)	(1.971)	(4.53)	(3.60)	(5.50)
Lagged Gini-squared (G^2_{t-1})	-2.318	-1.187	-4.825	3.79	-2.44	5.60
	(1.837)	(2.222)	(3.674)	(9.29)	(6.36)	(9.62)
Test for joint significance of G_{t-1} , G^2_{t-1} (p-value)	2.00	1.79	1.84	4.67	1.28	0.34
	(0.367)	(0.170)	(0.399)	(0.10)	(0.528)	(0.84)

Notes:
1. All specifications include the standard covariates (lagged crop share, market-quota price ratio, and log household education), plus year and province effects.
2. The IV specifications vary by instrument set. In each IV specification, the endogenous regressors are lagged income, the lagged Gini, and its square. Columns (3) and (4) use the level and Gini of per capita consumption, as well as the squared Gini as identifying instruments; Columns (5) and (6) use the "extended" instrument set described in Table 5.
3. The test for joint significance of the two quadratic terms is reported as a Chi-square test, except Column (2), which is an F-test.

Table 7: Running the regression "backwards" -- The Kuznet's relationship
Panel Data Specification, Fixed and Random Effects
Dependent variable: Change in Village inequality (Gini), ΔG_{it}
All specifications include period and province effects, and standard covariates
(standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	No IV		IV (1)		IV (2)		IV (3)	
Endogenous Regressor(s):	None		Lagged Gini (G_{t-1})		Lagged Gini (G_{t-1}) Growth (Δy_t)		Lagged Gini (G_{t-1}) Growth (Δy_t)	
Instrument set:	N/A		Consumption set		Consumption set		"Extended" set	
Fixed or Random Effects:	RE	FE	RE	FE	RE	FE	RE	FE
Lagged Gini (G_{t-1})	-0.321* (0.014)	-1.043* (0.074)	-0.232* (0.087)	-1.004* (0.178)	-0.225* (0.088)	-0.911* (0.201)	-0.281* (0.132)	-0.818* (0.184)
Contemporaneous growth (Δy_t)	-0.003 (0.014)	0.001 (0.014)	-0.003 (0.014)	0.001 (0.014)	-0.019 (0.033)	-0.051 (0.033)	-0.104 (0.085)	-0.092 (0.110)

Notes:
1. All specifications include the standard covariates (lagged crop share, market-quota price ratio, and log household education), plus year and province effects.
2. The four classes of specification (times FE and RE) vary the set of endogenous regressors, and the instrument set. The "consumption set" is the lag of the consumption gini and the lagged level of consumption (for columns (3) and (4)). Columns (5) and (6) use the same consumption instruments, but add the growth in consumption. Columns (7) and (8) use the "Extended" set of instruments described in Table 5.

Table 8: Cross-section growth regressions
Varying the specification and instrument set
Dependent variable: Village income from period 1 to 4
(robust standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS Estimates			IV: Consumption-based instruments		IV: Extended Set instruments	
Initial Income (y_0)	-0.022* (0.008)	-0.029* (0.008)	-0.041* (0.009)	-0.036* (0.011)	-0.060* (0.015)	-0.042* (0.011)	-0.045* (0.013)
Initial Gini (G_0)	-0.058 (0.047)	-0.099* (0.039)	-0.112* (0.045)	-0.125* (0.067)	-0.185 (0.101)	-0.167* (0.081)	-0.258* (0.111)
Initial Crop Share		-0.040* (0.013)	-0.031* (0.013)	-0.046* (0.019)	-0.045* (0.022)	-0.054* (0.018)	-0.051* (0.019)
Initial Market-Quota Price ratio		-0.007* (0.003)	-0.009 (0.006)	-0.006 (0.004)	-0.010 (0.006)	-0.005 (0.004)	-0.011 (0.006)
Initial (Log) Household education		0.027* (0.012)	0.042* (0.013)	0.031* (0.013)	0.050* (0.015)	0.034* (0.014)	0.037* (0.019)
First-stage F-test (income)				27.83* (0.0000)	13.86* (0.0000)	8.94 (0.0000)	9.76 (0.0000)
First-stage F-test (Gini)				22.47* (0.0000)	9.18* (0.0002)	6.86* (0.0000)	2.36 (0.030)
Hausman Test (IV vs. OLS) χ^2				1.03 (0.96)	0.99 (0.96)	-11.33 (N/A)	4.01 (0.55)
Overid test (P-value)						3.19 (0.67)	6.74 (0.240)
Province Effects?	No	No	Yes	No	Yes	No	Yes

Notes:
1. All standard errors are heteroskedasticity consistent.
2. The consumption-based instruments are the log of initial per capita consumption and the initial consumption gini. The extended set of instruments includes per capita cultivated land, per capita productive assets, the % of assets owned by the collective, the village dependency ratio (sum of young plus old divided by village population), the village coefficient of variation for the dependency ratio, and the village gini coefficients for household per capita land and household per capita productive assets. These are analogous to the instruments described in Table 5 (except for the cross-section)

Table 9: Reduced Form Regressions Corresponding to Table 8
Cross-section growth regressions
Dependent variable: Village growth
All specifications include province effects
(robust standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Initial Income		Initial Inequality		Growth in Incomes	
Initial Crop Share	-0.372* (0.156)		-0.043 (0.027)	-0.152* (0.044)	-0.015 (0.015)	-0.045 (0.030)
Initial Market-Quota Price ratio	-0.037 (0.038)		-0.018* (0.008)	-0.016* (0.008)	-0.004 (0.004)	-0.005 (0.005)
Initial (Log) Household education	0.192 (0.141)		-0.007 (0.023)	-0.024 (0.029)	0.039* (0.014)	0.026 (0.014)
Initial Consumption level	0.631* (0.125)		-0.011 (0.012)		-0.036* (0.008)	
Initial Consumption Gini	-1.286 (1.314)		1.206* (0.287)		-0.146 (0.122)	
Initial Land per capita		0.101 (0.078)		0.011 (0.012)		0.006 (0.009)
Initial Physical assets per capita		0.187* (0.037)		0.016 (0.009)		-0.013* (0.004)
Initial Percentage of Assets collective		0.055 (0.112)		-0.084* (0.027)		0.025* (0.012)
Initial Dependency Ratio		-2.350* (0.598)		0.184 (0.130)		0.117 (0.084)
Initial Gini, Land		0.527 (0.504)		0.087 (0.122)		-0.058 (0.060)
Initial Gini, Physical Assets		0.300 (0.203)		0.021 (0.050)		-0.041* (0.021)
Initial Coef Variation, Dep Ratio		-0.436 (0.288)		0.025 (0.073)		0.056 (0.041)
F-test for excluded instruments	13.86* (0.0000)	9.76* (0.0000)	9.18* (0.0002)	2.36* (0.030)	13.83* (0.0000)	3.38* (0.0032)
R-squared	0.752	0.774	0.630	0.591	0.390	0.414

Notes:
1. Columns 1, 3 and 5 are the reduced form (1^{st} stage) using consumption measures as instruments.
2. Columns 2, 4 and 6 are based on the extended set of instruments.
3. Sample size is 102 villages.
4. Standard errors are robust to heteroskedasticity.

Table 10: Potential Reconciliation of Panel and Cross-section results
Panel Data Specifications
Dependent variable: Village income growth
All specifications include period and province effects, plus standard covariates
(standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
	No Instruments		Instrument Set: Initial Consumption		Instruments Set: Initial Extended	
	RE	BE	RE	BE	RE	BE
Lagged Income (y_{t-1})	-0.208* (0.038)	-0.050 (0.039)	-0.258* (0.071)	-0.268* (0.080)	-0.097 (0.071)	-0.102 (0.070)
Lagged Gini (G_{t-1})	0.130 (0.205)	-0.175 (0.233)	-0.896 (0.661)	-0.895 (0.716)	-1.443* (0.719)	-1.302* (0.634)
Over-identification test (p-value)					4.68 (0.456)	18.79 (0.002)

Notes:
1. All specifications include the standard covariates, as described in Table 5, as well as period and province effects.
2. “BE” refers to the “Between Effects” estimator.
3. The instrumental variables specifications use the fixed, initial values of instruments, based on the consumption or “extended” instruments described in Table 5. The endogenous variables in the IV specifications are lagged income and the lagged Gini.
4. The over-identification test is reported only for the over-identified model (columns (5) and (6)).

Table 11: Cross-section estimates of income growth by major source
Does inequality affect the composition of income?
OLS and IV Estimates, including province effects
(robust standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	Agricultural Crop Income		Non-Ag Income		Wage Income		Family Business Income	
Initial Income (of that source) (y_{i0})	-0.030*	-0.031	-0.041*	-0.048*	-0.040*	-0.058*	-0.048*	-0.041*
	(0.014)	(0.021)	(0.008)	(0.011)	(0.007)	(0.012)	(0.010)	(0.020)
Initial Gini of total income (G_0)	0.224*	0.117	-0.215*	-0.493*	-0.417*	-0.450	0.097	-0.483
	(0.091)	(0.201)	(0.072)	(0.196)	(0.108)	(0.240)	(0.168)	(0.591)
Initial Education	0.029	0.025	0.061*	0.053	0.090*	0.109*	0.001	-0.025
	(0.023)	(0.029)	(0.026)	(0.029)	(0.031)	(0.037)	(0.034)	(0.039)
Initial crop share	0.130*	0.119	-0.060	-0.115*	-0.057	-0.107	-0.090*	-0.150*
	(0.041)	(0.068)	(0.032)	(0.047)	(0.039)	(0.062)	(0.045)	(0.056)
Initial Price Ratio	-0.001	-0.002	-0.013	-0.019*	-0.015	-0.014	-0.025	-0.035
	(0.009)	(0.010)	(0.007)	(0.009)	(0.009)	(0.011)	(0.017)	(0.021)

Notes:

1. Agricultural crop income is the total net income from grain and non-grain crops. Non-ag income is total net income earned from non-agricultural activity including wage income from hiring out, and net income from family run businesses. Wage income includes wages earned inside and outside the village.
2. Initial income in each regression is the initial income for *that particular source* (i.e., from agriculture, non-agricultural sources, etc.)
3. The IV versions of each regression use the extended set of instruments, described in the previous tables. The endogenous (instrumented) regressors are initial income, and initial inequality.

Table 12: Running the Regression in Reverse (Kuznet's Relationship)
Cross-section specification
Dependent variable: Change in Inequality
(robust standard errors in parenthesis)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS		Median Regression		IV (Cons Instruments)	
Initial Inequality (G_0)	-0.026* (0.008)	-0.030* (0.009)	-0.026* (0.012)	-0.030* (0.015)	-0.026* (0.013)	-0.035 (0.021)
Income Growth (Δy)	0.012 (0.019)	0.010 (0.019)	0.019 (0.027)	0.004 (0.020)	0.017 (0.019)	0.008 (0.019)
Initial Education	0.004 (0.002)	0.006* (0.003)	0.005 (0.004)	0.007 (0.004)	0.004 (0.002)	0.005 (0.003)
Initial Agriculture	0.011* (0.003)	0.008* (0.003)	0.015* (0.005)	0.011* (0.005)	0.011* (0.003)	0.008* (0.004)
Initial Market-quota Price Ratio	0.001 (0.001)	0.0004 (0.0012)	-0.0000 (0.0016)	-0.0002 (0.0018)	0.001 (0.001)	0.0003 (0.0012)
Province Effects	No	Yes	No	Yes	No	Yes

Notes:

1. The dependent variable is the change in village inequality over the sample period. The regression is similar to that in Table 8, except that the dependent variable is the change in inequality, and the key regressor is growth of incomes.
2. The specification is estimated by OLS (with and without province effects); as a median regression (with and without province effects), and by 2SLS, instrumenting initial inequality with the consumption instruments (initial consumption level and gini).
3. The OLS and 2SLS specifications report robust standard errors, while the median regression employs bootstrapped standard errors.
4. sample size is 102 villages.

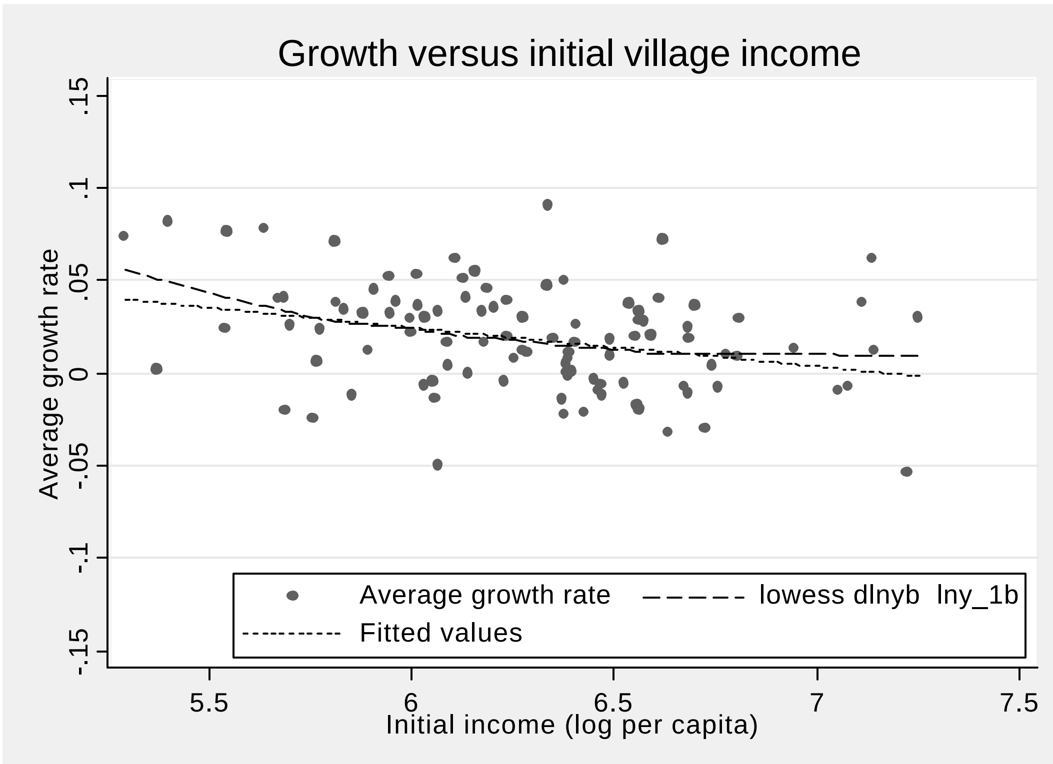
Table 13: Sensitivity to Definition of "Long Run Growth"
Using a shorter time period (cross-section specification)
(robust standard errors in parentheses)

	(1)	(2)	(3)	(4)
	Growth Income	Growth Crop Income	Growth Non-Agr. Income	Change in Inequality
Initial Inequality (G_0)	-0.052* (0.010)	0.210* (0.099)	-0.306* (0.091)	-0.061* (0.015)
Initial Income (y_0)	-0.068* (0.009)	-0.052* (0.014)	-0.071* (0.020)	
Growth (Δy)				0.045* (0.024)

Note:

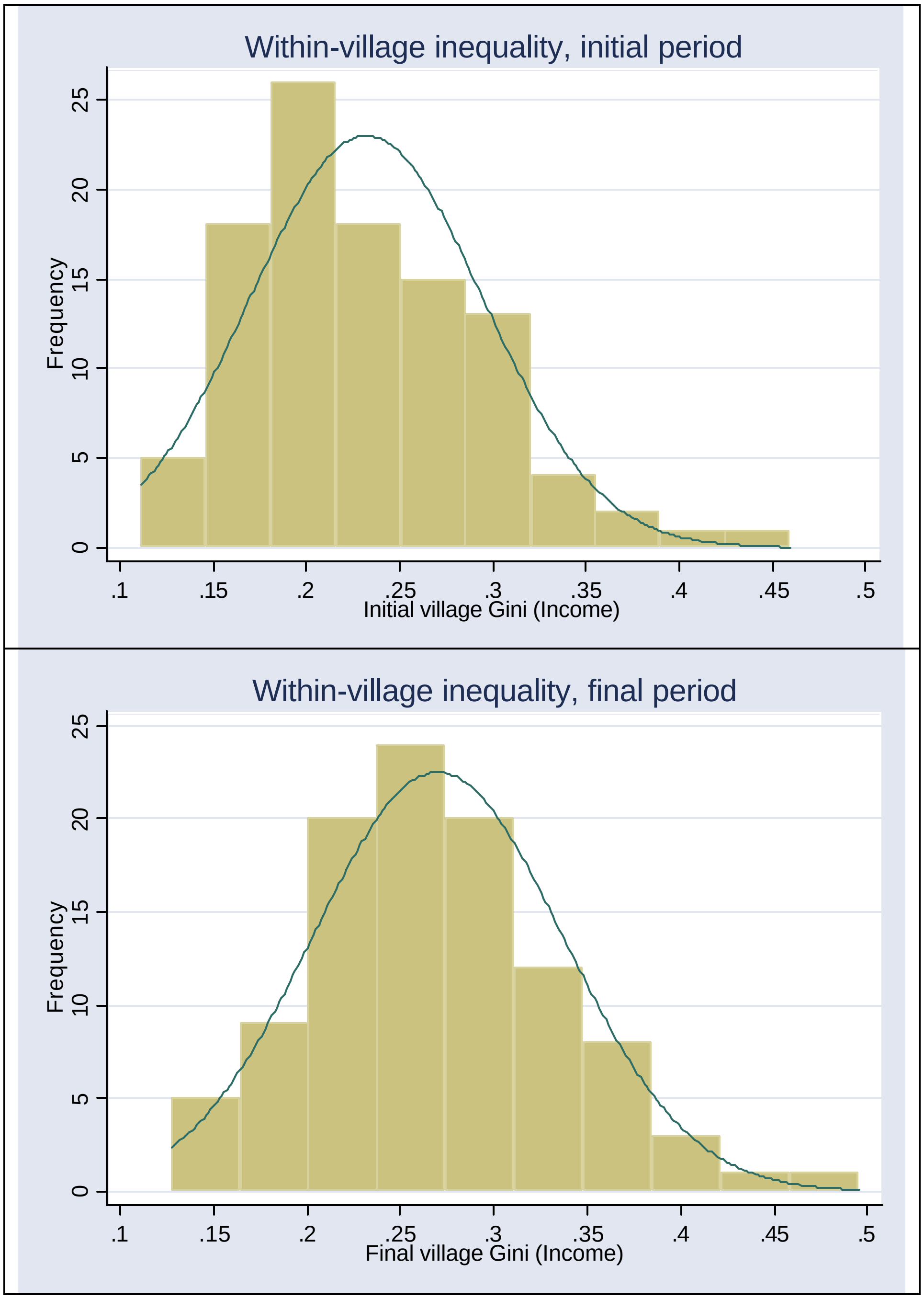
1. These specifications replicate the OLS regressions from the previous cross-section tables, but with period 3 taken as the end-point of the sample (instead of period 4).
2. In columns 1 and 4, initial income refers to initial total income. In columns 2 and 3, initial income refers to initial crop and non-agricultural income, respectively.
3. All specifications are estimated with provincial effects using OLS, and including the standard regressors (not shown).

Figure 1: Is there convergence in village per capita income?



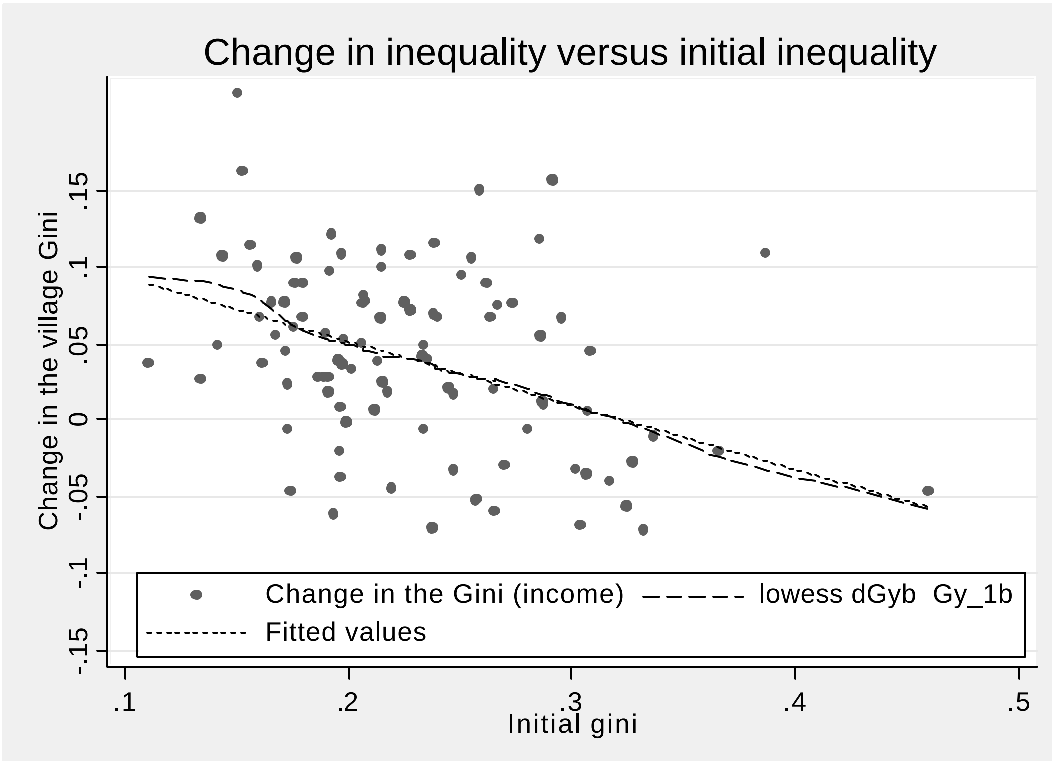
Notes: This figure shows the average growth rate of (real per capita) village income between Periods 1 and 4 (1986 to 1999) as a function of initial income. In addition to the scatter plot, we show an ordinary least squares, and lowess regression through the points. The coefficient of the slope in the linear regression is -0.021 (t=3.32).

Figure 2: The distribution of village Gini's in the initial and final periods



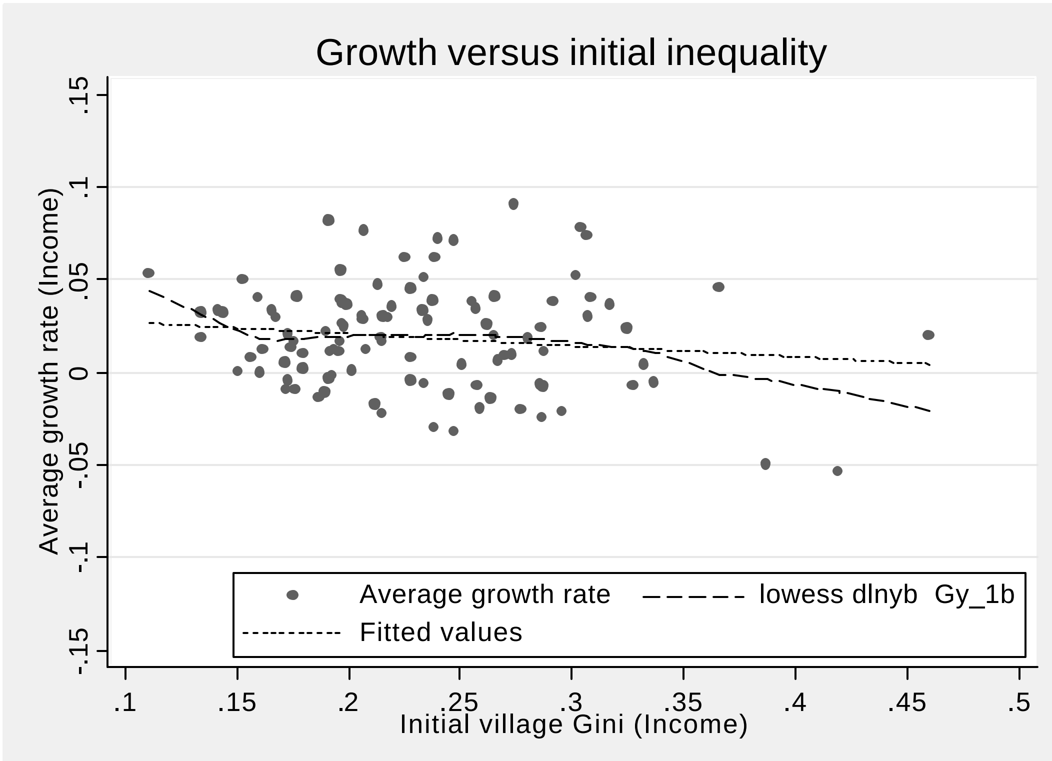
Notes: This figure shows histograms of the distribution of village-level Gini coefficients in the initial and final periods. We also show a super-imposed normal distribution.

Figure 3: Is there convergence in the level of income inequality?



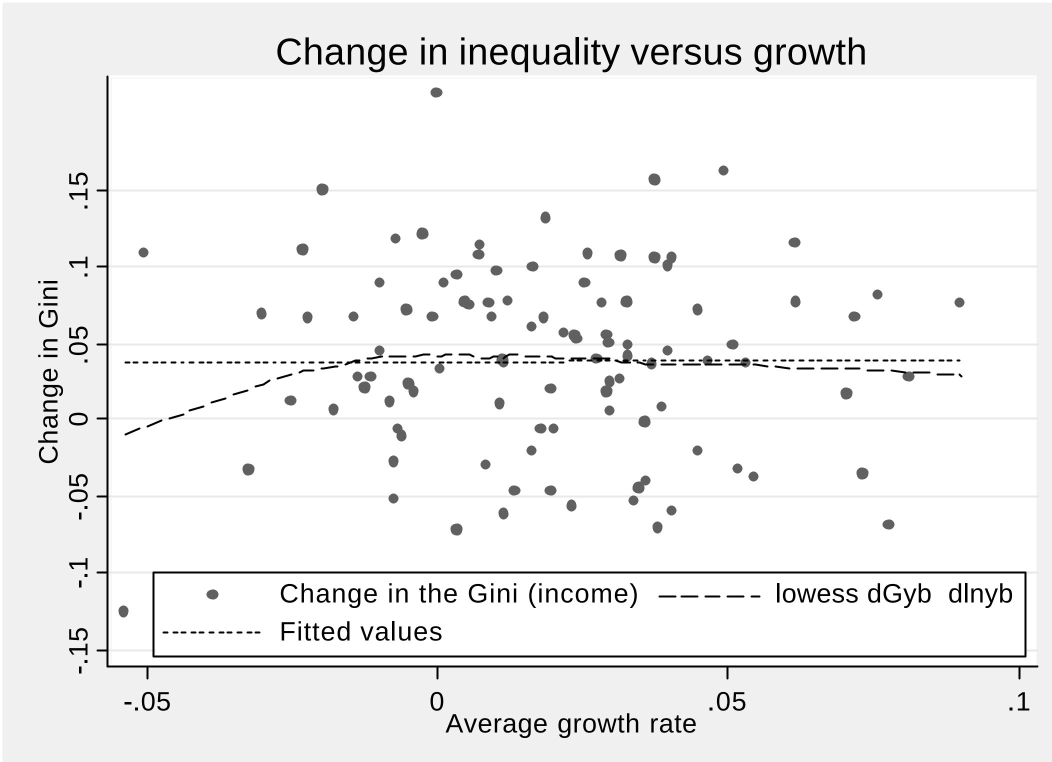
Notes: This figure shows the change in the village-level income Gini between 1986 and 1999, as a function of initial income inequality. In addition to the scatter plot, we show an ordinary least squares, and lowess regression through the points. The coefficient of the slope in the linear regression is -0.417 (t=4.62).

Figure 4: The relationship between initial inequality and subsequent growth (Income-based)



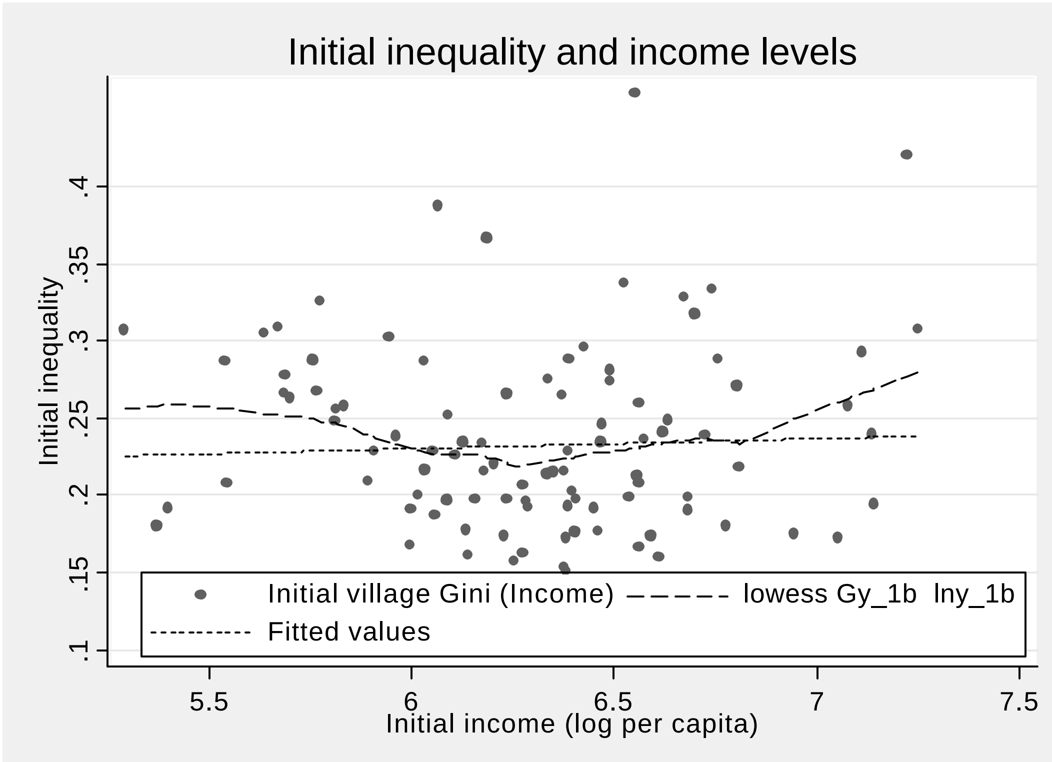
Notes: This figure shows the average growth of per capita village income between 1986-1999 as a function of initial income inequality. In addition to the scatter plot, we show an ordinary least squares, and lowess regression through the points. The coefficient of the slope in the linear regression is -0.063 (t=-1.42).

Figure 5: The relationship between *changes* in inequality and growth



Notes: This figure shows the average growth of per capita village income between 1986-1999 related to the change in the village income Gini. In addition to the scatter plot, we show an ordinary least squares, and lowess regression through the points. The coefficient of the slope in the linear regression is 0.017 (t=0.08).

Figure 6: The relationship initial inequality and initial income



Notes: This figure shows the initial Gini for income, plotted against the initial level of household per capita income. In addition to the scatter plot, we show an ordinary least squares, and lowess regression through the points. The coefficient of the slope in the linear regression is 0.06 (t=0.44).