GLOBALIZATION AND INEQUALITY:
EVIDENCE FROM WITHIN CHINA

Shang-Jin Wei
Yi Wu

Globalization and Inequality: Evidence from Within China
Shang-Jin Wei and Yi Wu
NBER Working Paper No. 8611
November 2001
JEL No. F1, O1

ABSTRACT

In this paper, we provide a case study of the impact of globalization on income inequality using data across Chinese regions. The literature on cross-country studies has been criticized because differences in legal systems and other institutions across countries are difficult to control for, and the inequality data across countries may not be compatible. An in-depth case study of a particular country's experience can provide a useful complement to cross-country regressions. We construct a measure of urban-rural income ratio for 100 or so Chinese cities (urban areas and adjacent rural counties) over the period 1988-1993. The central finding is that cities that experience a greater degree of openness in trade also tend to demonstrate a greater decline in urban-rural income inequality. Thus, globalization has helped to reduce, rather than increase, the urban-rural income inequality. This pattern in the data suggests that inferences based solely on China's national aggregate figures (overall openness and overall inequality) can be misleading. The negative association between openness and inequality holds up when we apply a geography-based instrumental variable approach to correct for possible endogeneity of a region's trade openness.

Shang-Jin Wei
International Monetary Fund
Room 10-700
19th Street, NW
Washington, DC 20433
and NBER
swei@imf.org

Yi Wu
Department of Economics
Georgetown University
Washington, DC 20057
wuy@georgetown.edu

"[C]ross-country regressions…are not the best tools for analyzing…the linkage between trade and growth." "[T]he most compelling evidence on this issue can come only from careful case studies…"

T.N. Srinivasan and Jagdish Bhagwati (1999, p9, and the abstract)

1. Introduction

This paper presents a case study of the impact of globalization on income disparity using data on Chinese cities.

There exists an active empirical literature on economic growth and income inequality, and a related and equally active literature on openness and economic growth. Most of the papers in these two literatures employ cross-country regressions. Prominent examples include Dollar (1992), Edwards (1992, 1998), Sachs and Warner (1995), Frankel and Romer (1999), Forbes (2000), Rodriguez and Rodrik (2000), Dollar and Kraay (2001a and 2001b) and other papers cited therein. Useful insights have been gained from this literature. However, analyses based on cross-country regressions have been criticized on two grounds.

First, data on the key variables under study may not be compatible internationally. This is particularly acute when it comes to income inequality: the definition and data collection method can be different across countries. For OECD countries, Atkinson and Brandolini (2001) noted the pitfalls in making cross-country comparisons based on pooled data. They observe that for a few countries where multiple measures of income distribution are available (households versus individuals, income versus consumption, etc.), the different measures can give different, sometimes contradictory, patterns even for the same time periods. Since the data that cross-country regressions have to rely on come from potentially different methodologies, they can produce misleading results when pooled together. The authors noted further that "in cross-country analysis, use of a dummy variable adjustment for data differences is not appropriate." It seems reasonable to assume that the data quality for developing countries is generally inferior to the OECD countries. Therefore, running cross-country regressions involving data from developing countries can only make the quality of inference worse.

Aside from the Atkinson-Brandolini criticism just mentioned, we should note another potential source of data incomparability. The validity of comparing living standards across countries depends on the validity of the so-called purchasing-power-parity adjustment, which in turn depends on the assumption that a common "representative consumption basket" can be meaningfully constructed for all countries. The last assumption cannot always be taken for granted.

Second, it is generally very difficult to control for differences in cultures, legal systems, or other institutions that may be relevant for the outcome variable under study (e.g., economic growth or income inequality). Inclusion of fixed effects in panel regressions helps. However, the myriad of country-specific institutions may also interact with the key regressor under investigation (e.g., openness) to affect the outcome variable (e.g., income inequality). In this case, the usual fixed effects are not sufficient to control for the influence of the country-specific institutions.

In an influential paper, T.N. Srinivasan and Jagdish Bhagwati (1999) asserted that cross-country regressions are deficient and cannot be relied upon to understand the impact of globalization on economic growth and presumably on inequality or other variables (see the quote at the beginning of the paper). One may not agree completely with Srinivasan and Bhagwati (1999). Nonetheless, given these criticisms, a careful case study of an individual country experience can, at a minimum, provide a useful complement to the literature based on cross-country regressions. Within a given country and over a relatively short time period, the culture, the legal system or other institutions can more plausibly be held constant. So the researchers' ability to isolate the effect of openness is enhanced. Furthermore, the comparability of data definition and collection method is, in principle, also higher within a single country than across multiple countries.

We choose China as the focus of our study for several reasons. First, the recent history of China offers a quasi-natural experiment on the consequence of embracing globalization. Before 1978, the country had relatively little trade with the rest of the world. In 1978, Deng Xiaoping-led Chinese government formally adopted "opening-to-the-outside-world" as a national economic policy. Since then, the trade-to-GDP ratio quadrupled (from 8.5% in 1977 to 36.5%

in 1999[1]).   When the central government reduces the policy barriers to trade, which includes policy-induced dis-incentives for exports as well as tariffs and non-tariff barriers to imports, all regions in China benefit. However, due to unequal natural barriers to trade (i.e., distance and access to the major seaports), the actual increase in openness varies widely across the different regions in China. This variation across space provides a good opportunity to study the impact of openness on inequality (and potentially other issues as well) while holding constant the legal system, macroeconomic policies, culture and a host of other factors.

Second, China is a large country, which implies a relatively large number of intra-national observations (e.g., there were 400 or so cities in China in the 1980s) that is convenient for a statistical analysis. For this reason, China provides better material for an in-depth case study than, say, Argentina, Bangladesh or Costa Rica, whose trade-to-GDP ratios also rose dramatically during the same period, but whose territories are much smaller.

Third, China's geography also turns out to be convenient for the type of statistical analysis that this paper carries out. An important issue in this research is that openness (or increases in openness) may be endogenous. In the literature, Frankel and Romer (1999) pioneered the technique of using geography as an instrument for openness: Geography is an important determinant of trade, and is arguably exogenous with respect to economic growth or income inequality. It would be desirable to adopt a similar instrumental variable strategy for single-country case studies. China is semi-landlocked. It has a coast on the East and Southeast sides, but is surrounded on other sides by tall mountains, deserts, or foreign territories that are minor participants in international trade. As a first order approximation, the differences across Chinese cities in terms of participation in international trade are, to a large degree, due to their varying distance from a major seaport. In fact, two seaports alone - Hong Kong and Shanghai - handled approximately 50% of China's total trade with the rest of the world during our sample period. This makes China a suitable case to apply the Frankel-Romer technique: we can construct an instrumental variable for a city's openness based on its access to major seaports. Such an instrumental variable would be much more difficult to construct for countries like the United States or Indonesia whose access points to international trade are more diffused.

---

[1] Trade data are from IMF's International Financial Statistics, various issues.

Due to data limitation, we focus on inequality between urban and rural residents rather than the more conventional measure of inequality in the distribution of income or wealth (as captured by Theil index or Gini coefficient, for example). However, the two concepts of inequality are not disconnected. The poor in China are disproportionately found in the rural areas. Conceptually, the overall income inequality (as measured by the Theil index, for example) can be linearly decomposed into three components: inequality within the urban areas, inequality within the rural areas, and inequality between the urban and rural areas. A number of authors have concluded that in the Chinese case, the inequality between the urban and rural areas dominates the inequality either within the urban areas or within the rural areas. The World Bank (1997) estimated that the urban-rural income inequality accounted for more than half of the overall income inequality in 1995, and the change in the urban-rural inequality explained about 75 percent of the change in the overall income inequality during 1984-1995.

Similar conclusions were reached in research papers that investigated a particular region or regions in China. For example, during 1986 to 1994, Yang (1999) estimated that the urban-rural inequality explained 82 percent of the change in the overall income inequality in Jiangsu province, and virtually all of the change in the overall inequality in Sichuan province. Yao and Zhu (1998) found that, in Sichuan and Liaoning provinces from 1988 to 1990, the urban-rural inequality accounted for between 47 to 51 percent of the overall income inequality. The contribution of the urban-rural inequality to the overall income inequality has been more or less stable over time, according to Kanbur and Zhang (1999), or has declined by a moderate amount, according to Khan and Riskin (1998). To summarize, while the exact estimates differ among the researchers, all agree that the urban-rural inequality is a major component of the overall income inequality in China.

Before we describe the actual patterns in the data, it is useful to note that the theoretical predictions with regard to the effects of openness on growth, and of growth on inequality, are ambiguous, depending on which theoretical model one uses. This point has been amply demonstrated by Rodrigues and Rodrik (2000) and Srinivasan and Bhagwati (1999). Given the ambiguity in the theories, the effect of openness on income inequality in China is a matter to be settled by the data. It is also useful to note at the onset that we would not necessarily insist that

the effect of openness in other countries has to be the same as in China. Again, institutions and other country-specific factors can matter here.

While the Chinese economy has dramatically increased its openness over the last two decades, the overall income inequality has risen as well. For example, Khan and Riskin (1998) estimated that the Gini coefficient in China increased from 38.2 in 1988 to 45.2 in 1995. With a different data set and methodology, the World Bank (1997) also estimated that the Gini coefficient in China had increased (from 28.8 in 1981 to 38.8 in 1995). From these aggregate statistics, it is tempting to conclude that embracing globalization has contributed to the rise in inequality. But this is not correct, as we will show later.

Our main findings can be summarized briefly here. Across China, openness and urban-rural inequality tend to be negatively associated. Those cities that have had a greater increase in the trade-to-GDP ratio have also tended to witness a reduction, rather than an increase, in the urban-rural income inequality. This pattern in the data continues to hold when we correct for possible endogeneity of a region's openness by adopting a geography-based instrumental variable approach.

Our data set does not allow us to investigate directly the exact channels through which globalization helps the rural residents. However, we can provide our hypothesis here. We note first a peculiar feature of the Chinese rural development. Due to the top leadership's concern about possible over-population in the cities, from the very start of the reform two decades ago, the government implemented the policy of "Li Tu Bu Li Xiang," or, the policy of "leaving-the-farm-work-but-not-the-farmland." What it means is that rural townships and villages are permitted to industrialize by converting some of the farmland to factories but farmers are discouraged from migrating to the cities. This policy is not executed perfectly as migration still occurs on a reduced scale, and has been relaxed a bit over time. However, as a result of the policy, there has been a spectacular rise in what are known as the "township-and-village enterprises," or TVEs, which are industrial entities located in the rural areas. Our presumption is that the rise of income in the rural areas reflects, to a large extent, the growth of the TVEs, though there has also been an improvement in the productivity of agriculture and in the terms of trade for agricultural products. Thus, globalization affects the Chinese rural areas by

accelerating the growth of the industrial firms in addition to affecting the agriculture sector directly.

In addition to the various papers cited above, there are a large number of research papers that look into either Chinese regional development or foreign trade performance. They are too numerous to cite them all. We note that Wei (1995) and Wang (1994) were the first two research papers that employed a city-level data set that is essentially the same one as in this paper (but in earlier years). Of the two, Wei (1995) examined the connection between openness and economic growth across the Chinese cities. Qian and Weingast (1996), and Jin, Qian, and Weingast (2001) investigated the role of fiscal decentralization in China's development, using a province-level data set. Young (2000) studied regional protectionism in China. Lardy (1992) is a classic reference on the Chinese trading regime during the reform era.

The paper is organized as follows. Section 2 describes the data and the construction of the key variables used in the analysis. Section 3 presents the statistical results and offers economic interpretations. Finally, Section 4 provides some concluding thought.

2. Data

The data used in this paper comes mainly from two sources: (1) <u>Urban Statistical Yearbook</u>, various issues, published by China's State Statistical Bureau, and (2) <u>Fifty Years of the Cities in New China: 1949-1998</u>, also published by the State Statistical Bureau[2].

The central variable under the investigation is the gap between urban and rural incomes. In order to explain the data set clearly, it is useful to provide a brief description of the Chinese administrative structure (see Figure 1). The entire country is divided into 27 provinces plus three province-status "super-cities" -- Beijing, Shanghai and Tianjin[3]. In each province (or super-city), the population is further divided into rural counties and cities. In many instances, cities are

---

[2] Most data come from the first source, except for GDP in 1992 and 1993, which are missing in the first source.

[3] The official term for super-cities are "directly administered cities" – meaning that the city officials report directly to the central government just as the officials in other provinces. Since 1997, Chongqing has been made a "super-city," bringing the total to four.

given the administrative power over the adjacent rural counties. For example, the municipal government of Shanghai administers 10 rural counties in addition to its urban area. The municipal government of Wuhan administers 3 rural counties. In 1994, 783 counties, or approximately 45% of all rural counties are administered by a total of 193 cities. The rest of the counties are not attached to any city (and administered mostly by prefecture governments, which are one level below provincial governments). Our data set consists of information on those cities that administer some adjacent rural counties. In other words, we exclude those counties that are unattached to any cities. We also leave out those (typically smaller) cities that are not authorized to administer any rural counties. In the rest of this paper, the term "city" is used to refer to an administrative unit with an urban area plus the adjacent rural counties under its jurisdiction.

For each variable (say, GDP or population) and each city (say, Wuhan), our data set provides information both at the level of a city (i.e., including the adjacent rural areas) and at the level of urban area of the city (i.e., excluding the rural areas). Let us use GDP(u, k, t) and POP(u, k, t) to denote GDP and population, respectively, for the urban part of city k in year t. Similarly, GDP(c, k, t) and POP(c, k, t) are used to denote GDP and population in year t for the entire city k that encompasses the adjacent rural areas. From this raw data, we can deduce the GDP and population for the associated rural areas alone, denoted by GDP(r, k, t) and POP (r, k, t), respectively. Obviously,

$$GDP(r, k, t) = GDP(c, k, t) - GDP(u, k, t)$$
and $$POP(r, k, t) = POP(c, k, t) - POP(u, k, t).$$

We measure the urban-rural income inequality in year t, Q(k, t), by the ratio of the two respective per capita incomes:

$$Q(k, t) = [GDP(u, k, t) / POP(u, k, t)] / [GDP(r, k, t) / POP(r, k, t)].$$

Our data source also reports values of exports by cities, but not information on imports dis-aggregated by cities. We define EXP(k, t) as the value of export in local currency by city k at year t. We then measure openness of city k in year t by

$$OPEN(k,t) = EXP(k, t)/GDP(k, t).$$

Note that the export data are available only at the city level (rather than separately for urban and rural areas). We omit the city subscript "c" in the defintion of openness.

In the subsequent statistical analysis, we focus on the change of this inequality measure from 1988 to 1993, $Q(k,93) – Q(k, 88)$. The beginning and ending points are constrained by data availability. 1988 was the first year when it became possible for us to compute $Q(k, t)$ for a sufficient number of cities. 1993 was the last year when $EXP(k, t)$ data were reported. As the reform deepened, an increasing number of corporations (including foreign-invested firms and some Chinese state-owned firms) earned the rights to conduct export and import business directly without having to go through the state-owned trading companies. A fraction of these exports did not get properly recorded at the city level. By 1994, the State Statistics Bureau made the judgement that the export data at the city level lost relevance due to this kind of reporting leakage and stopped reporting them altogether in its subsequent Urban Statistical Yearbooks. The consequence of this reporting leakage for our statistical analysis will be discussed later.

Before we implement the statistical analysis, we undertake a data cleanup. This includes (a) eliminating cities whose jurisdiction (e.g., number of counties under its administration) have changed over the sample period, and (b) correcting observations that appear clearly erroneous to us when checked against related series. The detail of the data cleanup is explained in an appendix. The final sample includes one hundred or so cities scatted around the country.

Table 1a reports the summary statistics for the key variables in the paper. A number of interesting observations can be made. First, at any given point in time and for a given variable, there is a tremendous amount of heterogeneity across space. For example, in 1988, the urban-rural income ratio, averaged across all the cities in the sample, was 2.87. But the standard deviation of this ratio was 1.17. The ratio spread from 1.09 for the most equal city to 7.33 for the most unequal city. Similarly, the export-to-GDP ratio in 1988 had a mean of 7.76 percent, but a standard deviation of 4.75. This spatial variation in openness and in inequality is necessary for us to conduct a meaningful statistical analysis. Second, over the six-year period from 1988 to 1993, while the average openness increased, the urban-rural income inequality had also gone up.

However, as we will show later, it would be misleading to conclude from these statistics that greater openness has contributed to greater inequality. Indeed, as suggested by the simple correlation reported in Table 1b, openness and inequality, across all cities in China, are negatively, rather than positively, correlated. We examine this more formally in the next section.

3. Statistical Analysis

In presenting the statistical results, we proceed in three steps. We start with a simple benchmark regression, and follow with a number of extensions and robustness checks. We then provide a discussion of the results.

Benchmark regression

Let $q(k, t) \equiv \log Q(k,t)$ denote the urban-rural inequality (in log) for city k in year t. We conceptualize that the inequality in a given city depends on a number of factors:

$$(1) \quad q(k,t) = \alpha + \beta \ open(k, t) + \gamma \ z(k) + h(k, t) \ \Theta + N(t) \ \Lambda + e(k, t)$$

This framework decomposes all the factors into four categories. (1) $open(k, t) \equiv \log[EXP(k, t)/GDP(k, t)]$ is the degree of trade openness (in log) for city k in year t. (2) $z(k)$ is a summary of city-specific factors that do not change over time. (3) $h(k,t)$ is a vector of factors other than $open(k, t)$ that are specific to city k and do change over time. Average income level of the city is one example in this category. (4) In addition to the city-specific factors, a number of national factors may affect the level of urban-rural inequality in all cities. Inflation rate and the terms of trade for agricultural products are two such examples. These factors are represented by the vector $N(t)$.

In Equation (1), $\alpha, \beta, \ \gamma, \ \Theta$ and Λ are parameters (or vectors of parameters of appropriate dimensions). The effect of embracing globalization on inequality is captured by the parameter β, whereas the overall level of inequality reflects a confluence of all of the factors in Equation (1). It is possible that other factors (say national inflation rate or national fiscal policy of

redistribution) can cause an increase in the overall urban-rural inequality even when the (partial) effect of globalization is to reduce the inequality (i.e., $\beta < 0$).

In our statistical analysis, we implement a first-differenced version of Equation (1).

(2) $Y(k) = \alpha' + \beta X(k) + H(k, t) \Theta + v(k)$

where $Y(k)$ denotes the change in inequality over the sample period,

(3) $Y(k) \equiv q(k, 1993) - q(k, 1988) = \log Q(k, 1993) - \log Q(k, 1988)$

$X(k)$ denote the change in the log trade openness for city k,

(4) $X(k) \equiv \log[OPEN(k, 1993)] - \log[OPEN(k, 1988)]$

and $\alpha' \equiv [N(1993)-N(1988)]\Lambda$, is a constant scalar. $v(k) \equiv e(k, 1993) - e(k, 1988)$, is a composite random variable, assumed to be iid normally distributed. $H(k) \equiv h(k, 1993) - h(k, 1988)$ is a vector of variables other than the change in openness that affect the change in inequality. The most noteworth feature of the first-difference specification in Equation (2) is that all factors that are common across the cities, and all factors that are specific to a city but invariant over time are eliminated by the differencing process.

Equation (2) is our benchmark specification. Regression results based on this specification are reported in Table 2. In Column 1 of this table, where the change in inequality is regressed on the change in openness alone, the slope coefficient is negative and statistically significant. This means that cities that experienced a bigger increase in openness, tend to witness a decline in urban-rural inequality. Figure 2 presents a scatter plot of the change in inequality against the change in openness across the Chinese cities. A negative association between these two variables is apparent from the graph, and is unlikely to disappear if we remove one or two observations.

Ravallion (2000) suggested the possibility that inequality may regress towards the mean over time. In Column 2, we add the initial inequality in 1988 as another regressor. The coefficient on the initial inequality is negative, suggesting a possible tendency for reversion: a city with a larger initial inequality tends to see a decline (or smaller increase) in subsequent inequality. However, this effect is not statistically significant. Note that after controlling for initial inequality, the coefficient on openness continues to be negative and statistically significant.

Perhaps economic growth may help to reduce the inequality because richer cities are more able to redistribute (implicitly as well as explicitly) to poor rural areas. In Column 3 of Table 2, the growth rate of city-level per capita GDP is included as another control. The coefficient on this new regressor is indeed negative, consistent with the "re-distributive ability" hypothesis, but it is only marginally significant (at the 15% level). One way to read this low significance level on the growth rate is that, holding openness constant, the per capita incomes in the rural and urban areas tend to grow in similar proportions[4].

One might be concerned that the results are driven by a few special coastal cities. Four cities were designated as "special economic zones." Fourteen coastal cities were designated as "coastal open cities." These cities were allowed to carry out certain types of market reform ahead of the rest of the country. While it is not clear how this would make inequality in these cities different from the other cities, it potentially can. In addition, if the investment in an urban area exceeds that in the adjacent rural areas (on a per capita basis), the gap between the urban and rural residents could widen. If the scale of the investment in the urban areas relative to the adjacent rural areas is correlated with that city's openness, omitting the relative investment measure could generate a spurious correlation between openness and inequality. In Column 4 of Table 2, we add two more control variables. The first is a dummy for cities that have been

---

[4] Using a cross-country regression framework, Dollar and Kraay (2001) documented that economic growth in terms of per capita GDP is largely uncorrelated with the share of the poor in total income. In their interpretation, therefore, economic growth is good for the poor as their income rises in lock steps with the average income of the country. Of course, as Ravallion (2000) pointed out, no change in the distribution implies that the rich gets a bigger share of the total gain from the growth.

designated as "Coastal Open Cities."[5] The second is the log ratio of investment in rural areas to urban areas[6]. In the regression, the dummy for "Coastal Open Cities" is not different from zero statistically. On the other hand, the ratio of investment in the urban relative to rural areas does matter as hypothesized: the coefficient is positive and statistically significant. Cities that have invested relatively more in the urban areas would see an increase in the urban-rural inequality. For our purpose, it is important to observe that after these controls, the coefficient on openness declines modestly in absolute value, but remains negative and statistically significant at the five percent level.

Correcting for Possible Endogeneity of the Openness Measure

A city's trade-to-GDP ratio may be endogenous. For example, the ratio may go up as a result of, rather than a cause for, a city's economic growth. Separately, the trade-to-GDP ratio may be mis-measured partly due to under-reporting of trade discussed earlier. To deal with this problem, we adopt a technique that was pioneered by Frankel and Romer (1999), and has subsequently been employed by Irwin and Tervio (2000) and Wei (2000). The basic idea in these papers is this: a country's volume of trade is related to its geography (e.g. proximity to other major trading nations in the world), but its geography is unlikely to be influenced by its income. In our case, we take advantage of the special geographic features of the Chinese territory to construct an instrumental variable for a city's openness, as noted in the introductory part of the paper. We observe that a different degree of openness reflects largely a different degree of access to major seaports. Furthermore, a small number of seaports carry out a large proportion of freight traffic. In fact, Hong Kong and Shanghai are by far the biggest ports for international trade in China. Together, about half of China's external trade pass through these two ports.

With these observations in mind, we use the distance from a city to either Shanghai or Hong Kong, whichever is smaller, as the instrumental variable (together with other regressors in

---

[5] The four "special economic zones" are not part of the sample due to definitional change of urban versus rural areas in these cities.

[6] Data on fixed capital investment is not available for 1993. We use the 1994 data instead.

the main regression as is standard with the two-stage least square approach) for openness for that city. By distance, we mean the greater circle distance between a pair of cities, computed by the so-called "oblique spherical triangle method" based on the latitudes and longitudes of the cities[7]. The information on the latitudes and longitudes of the Chinese cities is retrieved from the Defense Mapping Agency (1990).

To be more precise, suppose d(k, Shanghai) [or d(k, HongKong)] is the greater circle distance between city k and Shanghai (or Hong Kong), then, the key instrumental variable for city k is

(5) $D(k) = \min \{ \log[d(k, \text{Shanghai})] , \log[d(k, \text{HongKong})] \}$

Note that we use geography as an instrumental variable for the <u>change</u> in openness whereas Frankel and Romer (1990) use it for the <u>level</u> of openness. We justify our approach by noting the peculiar aspect of China's recent economic history. Until the end of the 1970s, China -- which means all its cities -- was very closed to the world trading system. The "opening-up" reform started by Deng Xiaoping in 1979 allowed the various cities to participate in international trade to an extent and in ways that had not been possible before. However, the increase in openness varies widely across the cities. We hypothesize that an important part of this difference is explained by the difference in their ability to access the major seaports. We will describe later the extent to which this hypothesis is true (reported in Table 4).

We replicate the regressions in Table 2 with a two-stage-least square approach. Shanghai and Shenzhen (which is next door to Hong Kong) are dropped from the regressions as we want to avoid the problem of having to define the distance for any of these two cities to itself[8]. The

---

[7] Oblique Spherical Triangle Method: Arc Distance $D = \text{Cos}^{-1}(\text{Sin(latitude1)} \times \text{Sin(latitude 2)} + \text{Cos(latitude 1)} \times \text{Cos(latitude 2)} \times \text{Cos(longitude1} - \text{longitude2)})$. Sign convention: + (-) for north (south) latitude, and +(-) for west (east) longitude. Distance in kilometers = 111.12*D.

[8] Shanghai and Shenzhen are also dropped in the sample in Table 2 so that the OLS and the 2SLS regressions have a comparable sample. Inclusion of these two cities in the OLS regressions makes no qualitative difference to the results.

results are reported in Table 3 (while the results from the first-stage regressions are reported in Table 4). Openness is negative and significant at the 5% level throughout the table. Figure 3 presents a conditional scatter plot of the change in the inequality against openness based on the IV regression in the last column of Table 3. As we can see, removal of any one or two observations on the chart is unlikely to change the negative slope between openness and inequality. If anything, the negative slope would be even steeper if one or two of the apparent outliers in the lower left corner of the graph were removed. We conclude, therefore, that greater openness helps to reduce the urban-rural inequality, and that this pattern is not a consequence of an endogenous trade-to-GDP ratio.

Note that the point estimates on openness are in fact bigger in the IV regressions than in the corresponding OLS regressions. While the reason for this is not immediately clear, we note that this pattern is similar to what Frankel and Romer found in their cross-country sample. If one takes the view that the slope estimates for openness in the OLS regressions in Table 2 are biased, then the IV estimates suggest that the impact of globalization on reducing income inequality is even bigger than what the simple OLS would suggest. According to the point estimate in Column 4 in Table 3, a ten percent increase in the trade-to-GDP ratio (e.g., from 0.2 to 0.22) leads to a three percent reduction in the gap between the urban-rural per capita incomes (e.g., from 5 to 4.85).

In the first stage of the 2SLS regressions just discussed, we have followed the usual practice in applied statistical work and employed all the regressors in the main regression other than openness as instrumental variables. However, one may be concerned that some of those regressors such as the growth rate of income are themselves susceptible to measurement errors. With this in mind, we also conducted 2SLS with $D(k)$ as the only regressor in the first-stage regression. The results (not reported to save space) are broadly similar to those in Table 3.

So far the key IV variable for a city's openness is defined as the minimum distance from the city in question to either Hong Kong or Shanghai. We note that while Hong Kong and Shanghai are the top two ports for China's international trade, they certainly do not cover all the trade. In particular, these two ports are on the East and South sides of the country. For cities in the north, the distance from a major seaport in the north may be a more relevant determinant for

their external trade. The biggest seaport in the northern segment of the coast is Qinghuadao. As a robustness check, we also define our key IV for city k's openness as the follows:

(6) $D^*(k) = \min \{ \log[d(k, \text{Qinghuadao})] , \log[d(k, \text{Shanghai})] , \log[d(k, \text{HongKong})] \}$

We have replicated the regressions in Tables 3 with this alternative instrumental variable, $D^*(k)$. The qualitative results are similar. In particular, the coefficient on openness is still negative and statistically different, although the point estimates tend to be somewhat smaller (regression results not reported to save space).

Other Robustness Checks and Extensions

As an alternative measure of a city's trade openness, we also use the trade-to-GDP ratio, averaged over the sample period. We replicate the key regressions in Tables 2 and 3, and report the results in Table 5. As one can see, the coefficients on this alternative measure are also negative and statistically significant for all regressions.

We use minimum distance to the two major seaports as an instrumental variable for openness. Note here, geography is used to instrument the level of trade openness, which is similar to the Frankel and Romer's original application. Figure 4 plots the average openness against the minimum distance to Shanghai or Hong Kong. The negative association between the two shows up very strongly. Going back to the main regression (Columns 5-8 in Table 5), we observe that the negative effect of openness on the urban-rural inequality is intact. In fact, the point estimates in the 2SLS regressions tend to be bigger (in absolute values) than the corresponding OLS estimates.

As a further check, we have collected data from additional years between 1988 and 1994 in order to form a panel. The advantage of a panel regression is its ability to make use of all available information on trade and inequality. A disadvantage is that the Frankel-Romer technique for an IV regression cannot be applied here since geographic features of a city (e.g., distance from any given city to Shanghai or Hong Kong) do not change over time. In any case, several different panel regressions are conducted and summarized in Table 6.

In Column 1 of Table 6, where city fixed effects are controlled for, the coefficient on openness is –0.04, which is smaller than the corresponding regression reported in Table 2. But the estimate is still negative and statistically significant. In the second column of Table 6, where city random effects are controlled for, the point estimate on openness is affected only slightly, remaining negative and significant. In the next two columns, we add year dummies in addition to the city fixed (or random) effects. The openness variable remains negative and statistically significant. The point estimates in the regressions in Columns 3 and 4 become bigger than the corresponding ones without the year dummies. In the last four columns of Table 6, we include an additional regressor, the investment in urban relative to rural areas. The coefficients on the openness variable are negative in all regressions. They are statistically significant at the ten percent level in the two fixed-effects specifications, but insignificant in the two random-effects specifications. A formal Hausman test rejects the null hypothesis that the coefficients in the fixed-effects and random-effects specifications are the same at the one percent level. This suggests that the random effects are correlated with the other regressors, and the random-effects specifications have produced biased estimates. Hence, the estimates in the fixed-effects specifications are more reliable. To sum up, the panel regressions also reveal the same pattern that openness is negatively associated with urban-rural inequality.

Further Discussion

We noted earlier that a city's level of trade may be underestimated due to reporting leakage. This was likely to be more serious at the end of the sample than at the beginning. In other words, the variable dopen(k, t) is measured with an error. The instrumental variable approach described earlier is one way to deal with this problem. The presence of measurement errors might be one reason for why the IV estimates for the openness effect tend to be somewhat bigger than the corresponding OLS estimates.

The left-hand-side variable, the urban-rural income inequality, may also be measured with error. If the measurement error is uncorrelated with the regressors (including openness), it would reduce the efficiency of the estimation but would not bias the estimates. However, unrecorded migration from rural to urban areas could induce a bias that is correlated with openness. We explain this in steps. First, rural migrants to an urban area are often not properly

recorded as part of the population in the urban area. Second, recorded city-level GDP is based on a product account. As a result, the output produced by rural migrants in a city may be recorded as part of the GDP for the urban area but not as part of the income of the rural area where the migrant comes from (Zhao, 1999). In this case, the recorded GDP per capita for an urban area may be too high relative to its true income per capita, whereas the recorded GDP per capita for a rural area may be too low relative to its true income per capita. This is likely to occur more at the end of the sample than at the beginning. Consequently, our measure of the change in urban-rural income inequality (as the change in the ratio of the two GDP per capita figures) may be upwardly biased. It seems plausible that this bias may be positively correlated with a city's change in openness (i.e. over time, more rural migrants go to cities that have a faster increase in the trade-to-GDP ratio). In other words, the true decline in the inequality may be bigger than what is recorded by the ratio of per capita GDP for cities that have a greater increase in openness. If that is the case, the estimated slope coefficient on openness may be too small (in absolute value) relative to its true value. Openness may have induced an even bigger reduction in the income inequality than our point estimates show.

4. Concluding Remarks

Using data on one hundred or so Chinese cities, the paper has documented a negative association between urban-rural inequality and openness: cities that have had a greater increase in the trade-to-GDP ratio also witness a faster reduction in the urban-rural income inequality. The instrumental variable approach strongly suggests that the direction of influence is from greater openness to less inequality.

This finding is in contrast with the impression one may obtain from the national aggregate figures: during the two decades of economic reform, both urban-rural inequality and trade openness have risen for the country as a whole. The paper suggests that it would be a mistake to conclude from the aggregate data that openness has contributed to a rise in inequality.

The case study in this paper provides a useful complement to studies based on cross-country regressions. However, the Chinese experience does not necessarily imply that the effect of openness on income inequality should be the same in other countries. Difference in institutions

and government policies could matter. It would be very useful to undertake similar case studies for other countries in the future so that the role of the institutions and policies in the impact of globalization can be understood better.

References:

Atkinson, Anthony B., and Andrea Brandolini, 2001, "Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries as a Case Study," <u>Journal of Economic Literature</u>, 39(3), September.

Cohen, B. Saul ed., 1998, <u>The Columbia Gazetteer of the World</u>, New York: Columbia University Press.

Defense Mapping Agency, 1990, <u>Gazetteer of the People's Republic of China,</u> Washington DC: Defense Mapping Agency.

Dollar, David, 1992, "Outward-oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976-85," <u>Economic Development and Cultural Change</u>, 523-544.

Dollar, David, and Aart Kraay, 2001a, "Growth is Good for the Poor," World Bank Policy Research Working Paper 2587.

Dollar, David, and Aart Kraay, 2001b, "Trade, Growth, and Poverty." unpublished World Bank Policy Research Working Paper 2615.

Economist, 2001, "Income Distribution In China," June 2, 39-44.

Edwards, Sebastian, 1992, "Trade Orientation, Distortions, and Growth in Developing Countries," <u>Journal of Development Economics</u> 39(1): 31-57.

Edwards, Sebastian, 1998, "Openness, Productivity and Growth: What Do We Really Know?" <u>Economic Journal</u> 108(March): 383-398.

Feder, Gershon, 1983, "On Exports and Economic Growth," <u>Journal of Development Economics</u>, 12(1/2): 59-73.

Forbes, Kristin, 2000, "A Reassessment of the Relationship Between Inequality and Growth," <u>American Economic Review</u> 90(4): 869-887, September.

Frankel, Jeffrey, A., and David Romer, 1999 "Does Trade Cause Growth?" <u>American Economic Review,</u> March, 379-399.

Hussain, Arthur, Peter Lanjouw, and Nicholas Stern, 1994, "Income Inequalities in China: Evidence from Household Survey Data," <u>World Development</u>, 22(12): 1947-57.

Irwin, Douglas A., and Marko Tervio, 2000 "Does Trade Raise Income? Evidence From The Twentieth Century," NBER working paper 7745, June.

Jin, Hehui, Yingyi Qian, and Barry R. Weingast, 2001, "Regional Decentralization and Fiscal Incentives: Federalism, Chinese Style," Unpublished working paper, Stanford University and University of California, Berkeley.

Kanbur, Ravi, and Xiaobo Zhang, 1999, "Which Regional Inequality? The Evolution of Rural–Urban and Inland-Coastal Inequality in China from 1983 to 1995," Journal of Comparative Economics, 27: 686-701.

Khan, Azizur R. and Carl Riskin, 1998, "Income Inequality in China: Composition, Distribution and Growth of Household Income, 1988-1995," China Quarterly June, 221-253

Khan, Azizur Rahman, Keith Griffin, Carl Riskin, and Zhao Renwei, 1993, "Sources of Income Inequality in Post-reform China," China Economic Review, 4(1): 19-35.

Knight, John, and Lina Song, 1999, "The Rural-Urban Divide: Economic Disparities and Interactions in China," New York: Oxford University Press.

Kwong, Tsz Man, 1994, "Markets and Urban-Rural Inequality in China," Social Science Quarterly, 75(4): 821-837.

Lardy, Nicolas R., 1992, "Foreign Trade and Economic Reform in China, 1978-1990," Cambridge: Cambridge University Press.

Lin, Justin Yifu, 1992, "Rural Reforms and Agricultural Growth in China," American Economic Review, 82(1): 34-51, March.

Lindert, H. Peter, and Jeffrey G. Williamson, 2001, "Does Globalization Make the World More Unequal?", NBER working paper 8228.

Munro, David, 1988, Chambers World Gazetteer – An A-Z of Geographical Information, Cambridge: Chambers.

Perkins Dwight and Shahid Yusuf, 1984, "Rural Development in China," Baltimore: Johns Hopkins University Press.

Qian, Yingyi, and Gerard Roland, 1998, "Federalism and the Soft Budget Constraint," American Economic Review, 88(5): 1593-1615, December.

Qian, Yingyi, and Barry R. Weingast, 1996, "China's Transition to Markets: Market-Preserving Federalism, Chinese Style," Journal of Policy Reform, 1: 149-185.

Ravallion, Martin, 2000, "Growth, Inequality, and Poverty: Looking beyond Averages," unpublished World Bank working paper.

Rodriguez, Francisco and Dani Rodrik, 2000, "Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence," NBER working paper No. 7081, April, in NBER Macroeconomics Annual 2000.

Rodrik, Dani, 1999, Making Openness Work: The New Global Economy and the Developing Countries, Washington, D.C.: The Overseas Development Council.

Rose, Andrew, 2000, "One Money, One Market: Estimating the Effect of Common Currencies on Trade," Economic Policy, April: 7-33.

Srinivasan, T.N., and Jagdish Bhagwati, 1999, "Outward-Orientation and Development: Are Revisionists Right?" Yale University, Economic Growth Center Discussion Paper No. 806.

Staiger, Douglas and Stock, James H., 1997, "Instrumental Variables Regression with Weak Instruments," Econometrica, 65(3): 557-86, May.

State Statistics Bureau, 1990, The Forty Years of Urban Development, Beijng: China Statistical Information and Consultancy Service Center.

State Statistics Bureau, 1993, China Statistical Yearbook, Beijng: China Statistics Publishing House.

State Statistics Bureau, 1995, China Statistical Yearbook, Beijng: China Statistics Publishing House.

State Statistics Bureau, Urban Statistical Yearbook of China, various years from 1985 to 1995, Beijing: China Statistics Publishing House.

State Statistics Bureau, 1999, Fifty Years of the Cities in New China: 1949-1998 (Xin Zhongguo Chengshi Wushi Nian), Beijing: Xinhua Press (in Chinese).

World Bank, 1994, China: Foreign Trade Reform, Washington, D.C.: World Bank.

World Bank, 1997, Sharing Rising Incomes – Disparities In China, Washington D.C.

World Bank, 1998, World Development Indicators, Washington D.C.

Welch, Finis, 1999, "In Defense of Inequality," American Economic Review, Papers and Proceedings, 89(2): 1-17, May.

Wang, Tao, 1994, "Market and Reform: The Growth of Chinese Urban Industries," unpublished, a chapter in Ph.D Dissertation, New York University.

Wei, Shang-Jin, 1995, "The Open Door Policy and China's Rapid Growth: Evidence from City-level Data," in Takatoshi Ito and Anne O. Krueger, eds., Growth Theories in Light of East Asian Experience, Chicago: University of Chicago Press.

Wei, Shang-Jin, 2000, "Natural Openness and Good Government," NBER Working Paper No. 7765, June.

Yang, Dennis Tao, 1999, "Urban-Biased Policies and Rising Income Inequality in China," American Economic Review Papers And Proceedings, 306-310, May.

Yao, Shujie, and Jirui Liu, 1998, "Economic Reforms and Regional Segmentation in Rural China," Regional Studies, 32(8): 735-746.

Yao, Shujie, and Liwei Zhu, 1998, "Understanding Income Inequality in China: A Multi-Angle Perspective," Economics of Planning 31: 133-50.

Young, Alwyn, 2000, "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China," Quarterly Journal of Economics, 115 (4): 1091-1135, November.

Zhao, Yaohui, 1999, "Labor Migration and Earnings Differences: the Case of Rural China," Economic Development and Cultural Change, 47(4): 767-782, July.

Figure 1: Administrative Structure in China (1993)

```
                    ┌─────────────────────────┐
                    │   Central Government     │
                    └─────────────────────────┘
                       /                    \
        ┌──────────────────┐         ┌──────────────────┐
        │  Provinces (27)  │         │ Super Cities (3) │
        └──────────────────┘         └──────────────────┘
          /            \                  /          \
  ┌─────────────┐  ┌─────────────────────────┐  ┌──────────┐  ┌──────────┐
  │ Prefectures │  │ Prefecture-level Cities │  │  Urban   │  │  Rural   │
  └─────────────┘  │         (184)           │  │  Area    │  │ Counties │
    /        \     └─────────────────────────┘  └──────────┘  └──────────┘
                        /            \
┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐
│  Rural   │ │Low-Level │ │  Urban   │ │  Rural   │
│ Counties │ │  Cities  │ │  Area    │ │ Counties │
│          │ │  (386)   │ │          │ │          │
└──────────┘ └──────────┘ └──────────┘ └──────────┘
```

Our Data Set
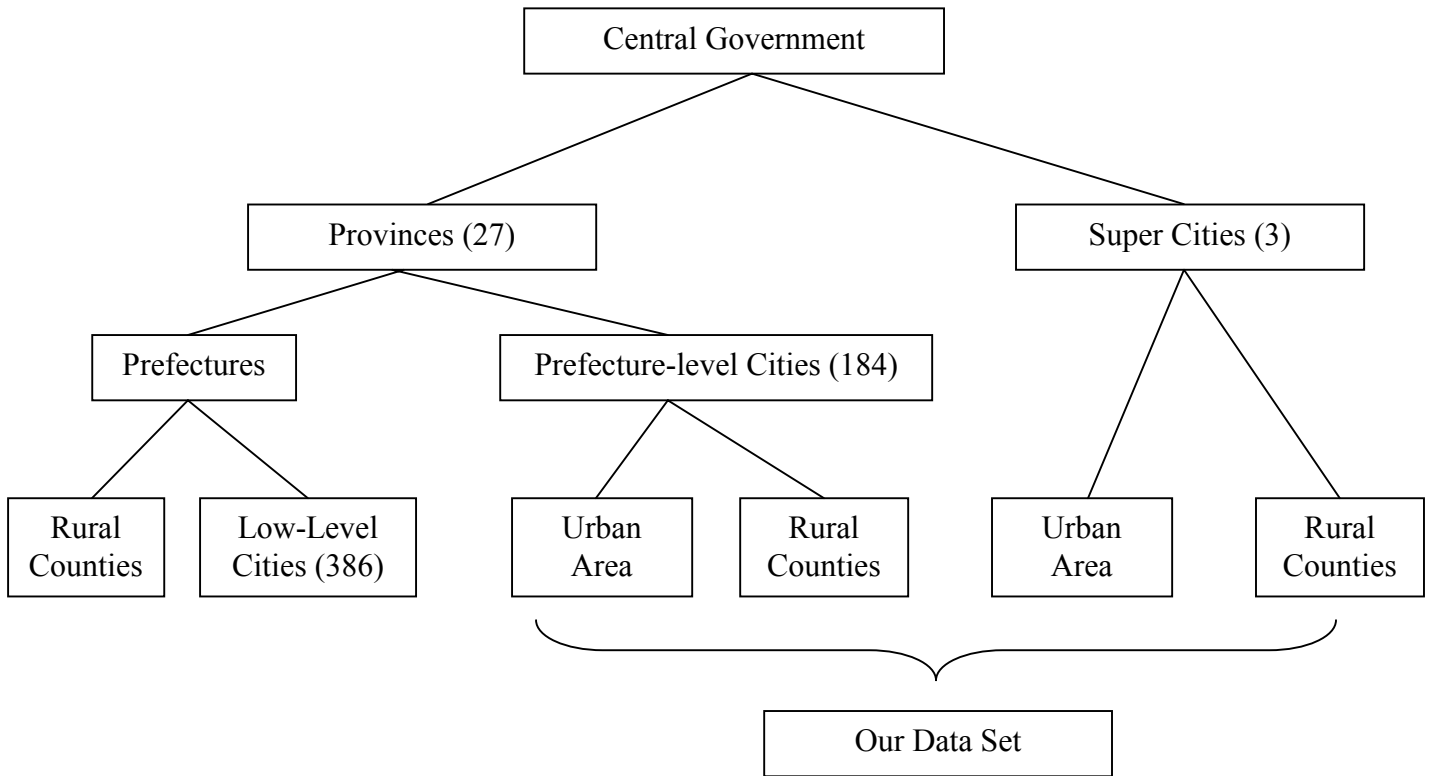
Note: In 1994, there were 783 counties that were administrated by cities, accounting for 45.1% of the total number of counties in China (1993 data not available).

Figure 2: Openness and Urban-Rural Income Disparity:
Simple Correlation
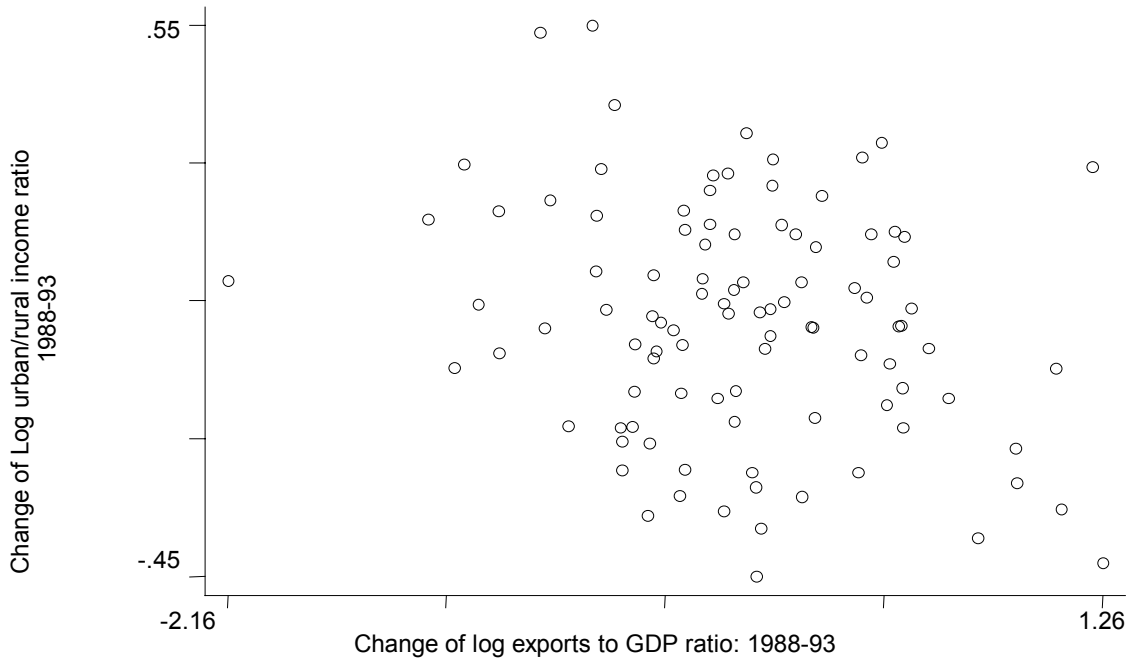
Figure 3: Openness and Urban/Rural Income Disparity:
Conditional Correlation from an IV Regression

coef = -.28, se = .10, t = -2.79

Figure 4: Openness and Minimum Distance to Shanghai or Hong Kong
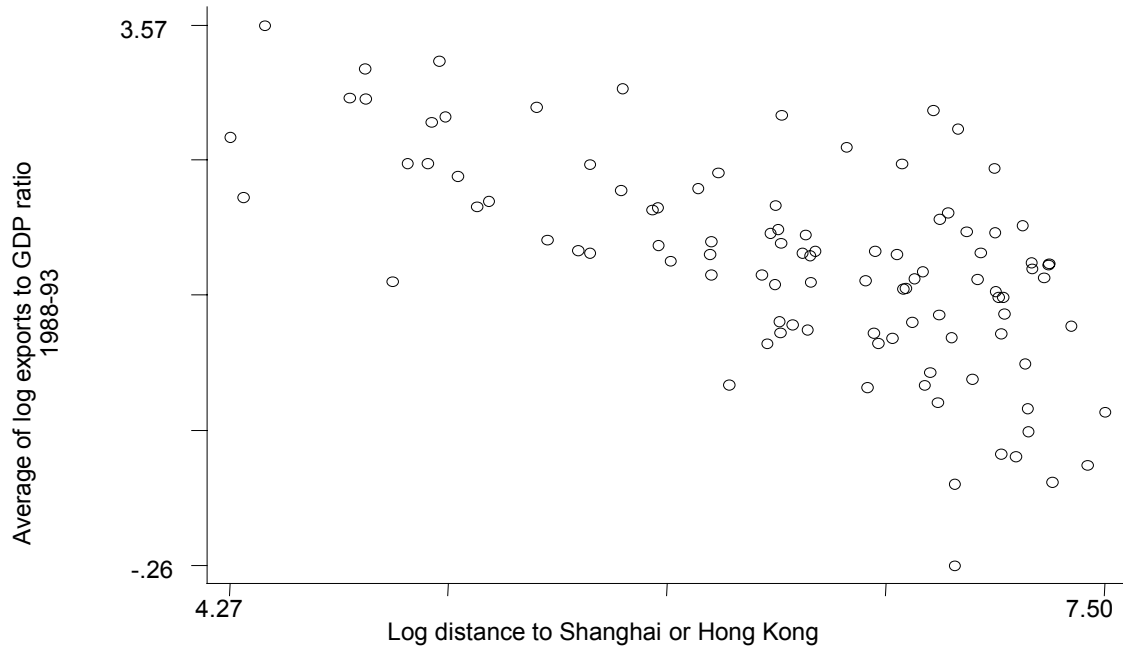
Table 1a: Summary Statistics

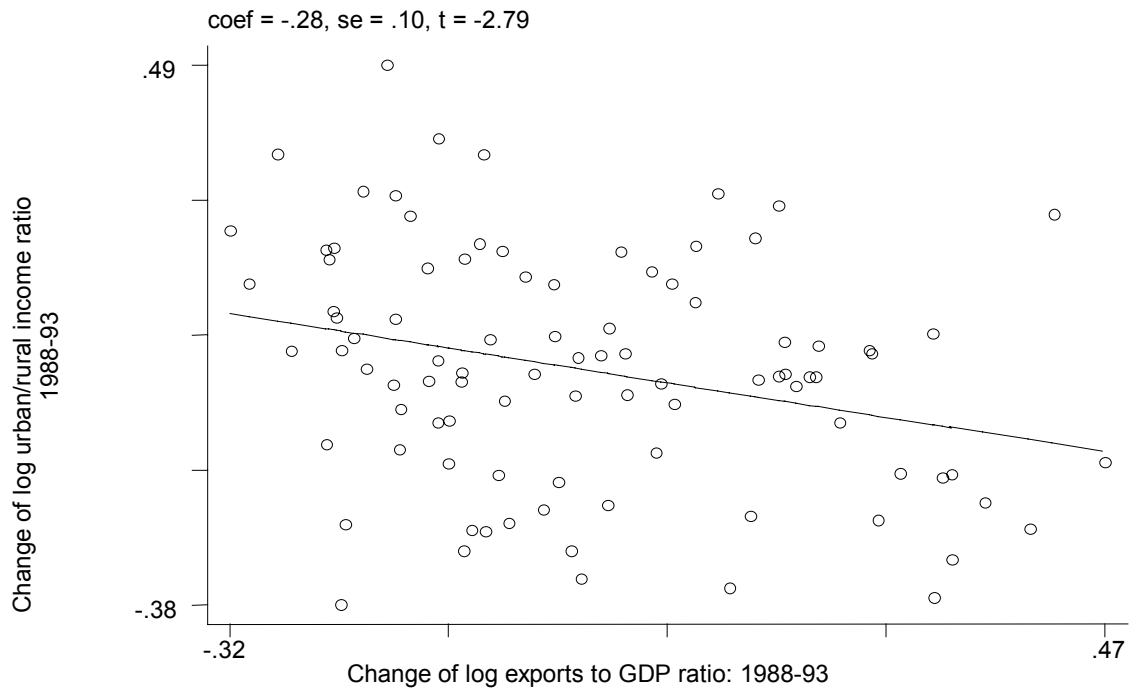| Variable	Obs	Mean	Std. Dev.	Min	Max
Per Capita GDP in 1988					
Urban GDP p.c.	100	2808	1240	663	7626
Rural GDP p.c.	100	1058	516	320	3276
Urban/Rural Income Ratio	*100*	*2.87*	*1.17*	*1.09*	*7.33*
Per Capita GDP in 1993					
Urban GDP p.c.	100	6307	3245	1355	18547
Rural GDP p.c.	100	2507	1724	413	10473
Urban/rural income ratio	*100*	*2.95*	*1.54*	*0.95*	*11.62*
Export-to-GDP Ratio					
1988	100	7.76	4.75	0.90	27.69
1993	100	8.56	10.30	0.43	76.39
Growth Rate of					
Per Capita GDP, 1988-93	253	6.61	5.67	-10.9	23.3

Note: All ratios are expressed in percentage term.

Table 1b: Pair-wise Correlation

	Change in urban/rural income ratio	Growth of per capita GDP	Average openness	Change in openness
Growth	-0.13			
Average openness	-0.22	0.06		
Change of openness	-0.11	0.09	0.65	
Log distance to Shanghai or HK	0.26	-0.24	-0.66	-0.37

Note: "Distance" refers to the minimum of direct distance from a city to either Shanghai or Hong Kong.

Table 2: Openness and Urban-Rural Income Inequality,
(OLS in First Difference with Robust Standard Errors)

Dependent variable: Change in log (urban GDP p.c./ rural GDP p.c.) from 1988 to 1993				
	(1)	(2)	(3)	(4)
Change in Log (exports / GDP) Over 1988-93	-0.083** (0.035)	-0.085** (0.035)	-0.092** (0.035)	-0.077** (0.034)
Initial Inequality in log		-0.031 (0.053)	-0.039 (0.053)	-0.048 (0.049)
Growth rate of p.c. GDP			-0.009# (0.006)	-0.007 (0.006)
Dummy for costal open cities				-0.018 (0.067)
Change in log ratio of urban/rural fixed capital per capita (88-94)				0.072* (0.037)
Adjusted R-squared	0.04	0.04	0.05	0.06
No. of Obs.	100	100	100	93

Notes: Robust standard errors are in parenthesis. **, *, and # denote statistically significant at the 5%, 10%, 15% levels, respectively. An intercept is included in all the regressions but not reported to save space.

Table 3: Instrumental Variable Regressions

Dependent variable: Change of log (urban GDP p.c./ rural GDP p.c.) from 1988 to 1993

	(1)	(2)	(3)	(4)
Methodology	IV	IV	IV	IV
Change in Log (exports / GDP)	-0.301**	-0.319**	-0.271**	-0.282**
During 1988-93	(0.128)	(0.133)	(0.112)	(0.122)
Initial inequality in log		-0.050	-0.057	-0.059
		(0.064)	(0.060)	(0.058)
Growth rate of p.c. GDP			-0.012*	-0.013*
			(0.007)	(0.008)
Dummy for costal open cities				-0.018
				(0.079)
Change in log ratio of urban/rural				0.076*
fixed capital p.c: 1988-94				(0.044)
No. of Obs.	100	100	100	93
First-stage F on the instruments	11.49	11.07	13.9	10.9
p-value for Hausman test	0.03	0.03	0.05	0.03

Notes:
1. R^2's in the IV regression are not reported as they do not have the standard interpretation.
2. The null hypothesis for the Hausman test is that the coefficients in the OLS and the IV regressions are not different systematically. A rejection of the null implies that the OLS estimate is biased.
3. First-stage F is the F-statistic for the null hypothesis that the coefficients on the instruments are zero.

Table 4: First-stage Regressions in 2SLS

Dependent variable: change in log (export/GDP) 1988-93				
Corresponding to the regressions in Table 3	(1)	(2)	(3)	(4)
Minimum Log distance to Shanghai or Hong Kong	-0.237***	-0.237***	-0.266***	-0.251***
	(0.070)	(0.071)	(0.072)	(0.076)
Initial inequality		0.002	-0.016	0.009
		(0.145)	(0.143)	(0.145)
Growth rates from 1988 to 1993			-0.032**	-0.045***
			(0.015)	(0.017)
Dummy for costal open cities				-0.066
				(0.198)
Change of log ratio of urban/rural fixed capital p.c.: 1988-94				-0.047
				(0.112)
Adjusted R-squared	0.10	0.09	0.12	0.10
No. of Obs.	100	100	100	93

Table 5: Alternative Measure of Openness

Dependent variable: Change of log (urban GDP p.c./ rural GDP p.c.) from 1988 to 1993								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Methodology	OLS	OLS	OLS	OLS	IV	IV	IV	IV
Log (exports / GDP)	-0.101***	-0.101***	-0.098***	-0.071**	-0.125***	-0.128***	-0.125***	-0.136***
Averaged over 1988-93	(0.028)	(0.028)	(0.029)	(0.031)	(0.044)	(0.043)	(0.045)	(0.051)
Initial inequality		-0.023	-0.026	-0.039		-0.023	-0.025	-0.035
		(0.051)	(0.051)	(0.049)		(0.051)	(0.052)	(0.050)
Growth rate of p.c. GDP			-0.003	-0.002			-0.002	-0.001
			(0.005)	(0.006)			(0.006)	(0.006)
Dummy for costal open cities				0.028				0.070
				(0.070)				(0.076)
Change in log ratio of urban/rural				0.082**				0.092**
fixed capital p.c.: 1988-94				(0.038)				(0.039)
Adjusted R-squared	0.11	0.10	0.10	0.06	.	.	.	.
No. of Obs.	100	100	100	93	100	100	100	93
First-stage F on instrument					68.38	71.55	66.16	54.69
p-value for Hausman test					0.47	0.40	0.44	0.09

Notes:
1. R^2's in the IV regression are not reported as they do not have the standard interpretation.
2. The null hypothesis for the Hausman test is that the coefficients in the OLS and the IV regressions are not different systematically. A rejection of the null implies that the OLS estimate is biased.
3. First-stage F is the F-statistic for the null hypothesis that the coefficients on the instruments are zero.

Table 6: Panel Regression

Dependent variable: log (urban GDP p.c. / rural GDP p.c.)

Methodology	(1) Fixed effects	(2) Random effects	(3) Fixed effects	(4) Random effects	(5) Fixed effects	(6) Random effects	(7) Fixed effects	(8) Random effects
Log (exports / GDP)	-0.041**	-0.038**	-0.047***	-0.040**	-0.055**	-0.035#	-0.042*	-0.029
	(0.018)	(0.017)	(0.018)	(0.017)	(0.024)	(0.022)	(0.025)	(0.022)
Log GDP p.c. for city and Counties	-0.041**	-0.038**	-0.198***	-0.122**	0.048	0.064*	0.166	0.136**
	(0.016)	(0.016)	(0.065)	(0.053)	(0.042)	(0.038)	(0.119)	(0.061)
Log urban/rural p.c. fixed capital					0.004	0.162***	0.016	0.168***
					(0.031)	(0.024)	(0.031)	(0.024)
Year dummy	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.02	0.002	0.05	0.002	0.02	0.45	0.04	0.45
No. of Obs.	666	666	666	666	397	397	397	397
p-value for Hausman test		0.71		0.83		0.00		0.00

Note: The null hypothesis for the Hausman test is that the coefficients in the fixed-effects and random-effects specifications are not different systematically. A rejection of the null implies that the random effects are correlated with the other regressors, and hence the estimates from the random-effects specification are biased.

Appendix A: List of cities in the sample, 1988-93

Anqing, Baiyin, Baoding, Baoji, Baotou, Beihai, Bengbu, Benxi, Changchun, Changzhi, Changzhou, Chengdu, Chifeng, Chongqing, Dalian, Deyang, Dongying, Foshan, Fushun, Fuxin, Fuzhou, Guangyuan, Guilin, Hangzhou, Hebi, Hefei, Hengyang, Hohhot, Huaibei, Huainan, Huaiyin, Huangshan, Huangshi, Huizhou, Huzhou, Jiaozuo, Jiaxing, Jilin, Jincheng, Jinhua, Jiujiang, Kaifeng, Leshan, Lianyungang, Liaoyang, Liaoyuan, Liupanshui, Liuzhou, Luohe, Luoyang, Maoming, Meizhou, Mianyang, Nanchang, Nanjing, Nantong, Ningbo, Panjin, Putian, Qingdao, Qinhuangdao, Qitaihe, Quzhou, Sanmenxia, Sanming, Shanwei, Shaoguan, Shaoxing, Shaoyang, Suzhou, Tangshan, Tianjin, Tianshui, Tongchuan, Tonghua, Tongling, Wuxi, Wuzhou, Xian, Xiangfan, Xianyang, Xinxiang, Xinyu, Xuchang, Yancheng, Yangquan, Yantai, Yinchuan, Yingkou, Yingtan, Yueyang, Zaozhuang, Zhangzhou, Zhanjiang, Zhaoqing, Zhengzhou, Zhenjiang, Zhoushan, Zhuzhou, Zigong

Appendix B: Data Cleaning

In three cases, observations are dropped to ensure that the quality of the data exceeds a minimum threshold.

a) Cites with at least one change in jurisdiction between 1990 and 1993 as listed in <u>Fifty Years of the Cities in New China: 1949-1998</u>. As an example, the number of rural counties under a city's administration might increase from one to four. Note that changes in jurisdiction prior to 1990 were not listed.

b) Cities whose urban or rural population had a change either by more than 40% or by more than 400,000 people in a single year. We suspect that these cities also experienced a change in their jurisdictions that was not properly recorded in the published sources.

c) Cities with obvious errors or major abnormality in one of its key variables. For example, a city may have a virtually constant population in two adjacent years (e.g., 1987 and 1988), but its GDP in later year (1988) is only ¼ of the previous year (1987).

In some rare instances when there is an obvious way to fix a data error, we do that. For example, a city's recorded ratio of the urban-rural GDP per capita in 1993 may be only half of both its 1992 and 1994 levels. There are six such cities in our sample. In this case, we use the average of the 1992 and 1994 levels to replace the recorded 1993 values. However, replacement of the recorded values is done relatively rarely. We generally choose to err on the conservative side: when there is no obvious and non – arbitrary way to fix a data error, we choose to drop the observation from the sample.