Who Gains from Which Infrastructure in China?

Xun Zhang
(Beijing Normal University)

Guanghua Wan
(Asian Development Bank Institute)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 6E: Globalization and Inclusive Growth

Time: Friday, August 24, 2018 [09:00-12:30]

Who Gains from Which Infrastructure in Rural China?

Xun Zhang[†] Guanghua Wan[$]

Abstract: The distributive impacts of infrastructure have largely been overlooked despite a growing literature on its role in promoting growth. This paper will: (i) demonstrate the deficiency of the conventional approach to modeling inequality; (ii) extend the Mincer earnings function so that both growth and distributive effects of infrastructure can be estimated; and (iii) fit the extended model to a large sample of individual-level data from rural China for the period of 1989–2011, providing estimates of growth and distributive impacts of specific physical infrastructures: telephone, tap water and electricity. All these infrastructures are found to promote rural income growth, helping narrow the rural–urban gap, which is the dominant component of China's overall inequality. Further, the poor are found to gain more than the rich, implying benign distributive effects of these infrastructures. In addition, males, the more experienced, the better educated, and to some extent the married benefited more than their counterparts, especially from telephone. Finally, some of these subpopulation effects became more significant in recent years and are larger in inland China. The empirical results are robust to different definitions of the experience variable, consideration of the mortality selection bias, reconstruction of the telephone data, and possible reverse causality.

Keywords: Infrastructure; Distributive effects; Inequality; Inclusive Growth; China

JEL: D31, H52, H54

[†] Beijing Normal University, China. Email: zhangxun@bnu.edu.cn.
[#] Asian Development Bank Institute, Japan. Email: gwan@adbi.org.

This paper is funded by Bairen Program of Yunnan province and the Project 71133004 and 71603026 supported by National Natural Science Foundation of China. It is also funded by Project 2015M580055 and 2016T90048 funded by China Postdoctoral Science Foundation.

1. Introduction

It has been established that infrastructure is an important and essential force driving productivity improvement and economic growth (Gramlich 1994). According to the World Bank (1994), infrastructure provides access to basic services, facilitates human/physical capital accumulation, promotes trade via linkages to markets, lowers production/transaction costs, and helps improve the environment. Moreover, infrastructure investment is known to directly generate jobs and may lead to inflows of investment to lagging or previously isolated areas, potentially producing beneficial distributive effects. In developing countries, connectivity infrastructure, such as roads and communication, facilitates migration and the emergence of rural non-farm activities. Both are important for bridging the rural–urban disparity that prevails in many economies (Shorrocks and Wan 2005).

While the literature on the growth effects of infrastructure is sizable and increasing (see the next section for more details), there is a shortage of research work on its distributive or inequality effects. This is regrettable as rising inequality has become a major socioeconomic issue in developing, emerging, as well as developed countries. If market forces do lead to rising inequality, as argued by Piketty (2014), government interventions become inevitable. Piketty (2014) suggests the imposition of taxes to help raise public revenue. But equally or more important is public spending. And indeed fast-growing economies spend a significant proportion of government budget on infrastructure. Even industrialized countries are confronted with infrastructure upgrading.[1] Clearly, the distributive impacts of infrastructure provision or spending, largely unaddressed in the literature with a few exceptions (see Section 2), can no longer be overlooked, particularly given the pursuit of inclusive growth by more and more institutions and governments all over the world.

China offers a natural setting to shed light on the distributive impacts of infrastructure. China's investment-driven growth model has been accompanied by huge investment into infrastructure. At the onset of economic reform that began in late 1978, infrastructure

[1] Further, up to 70% of lending by multilateral development banks such as the World Bank and Asian Development Bank is in infrastructure, let alone the forthcoming Asian Infrastructure Investment Bank.

investment only accounted for 5.44% of gross domestic product (GDP). This percentage more than tripled, reaching 18.19% in 2010. Note that this tripling was accompanied by fast growth in GDP at an annual rate of almost 10% throughout this period. As a consequence, rural infrastructure saw significant improvement. For example, the amount of investment in rural hydropower construction in 2010 was 12 times that in 1990. Rural consumption of electricity increased from almost zero in 1953 to 663 billion kilowatts in 2010. By then, 99% of villages in China had gained access to electricity and more than 98% of households in the villages had gained access to electricity. In terms of telecommunications, the number of landline telephone subscribers rose from 1.47 million in 1990 to 97.8 million in 2010.[2]

Meanwhile, worsening income distribution has been ranked among the top three most serious socioeconomic and policy issues for decades in China (Wan 2007, 2008a, 2008b; Wang, Wan, and Yang 2014). As Figure 1 illustrates, the overall regional inequality measured by the Theil index rose from a low level of less than 0.04 in 1983 to an alarmingly high 0.18 in 2009, almost quadrupling within a short period of two and half decades. Worse still, inequality within rural China increased faster than its urban counterpart. The much steeper slope of the overall inequality curve than either the rural or urban curve implies a large and growing urban–rural gap. This contrasts with the fact that China's massive infrastructure investment aims at narrowing down regional as well as urban–rural gaps. Clearly, it is important to analytically explore the distributive impacts of infrastructure in China.

[Insert Figure 1 approximately here]

This paper represents one of the first efforts to analyze the distributive impacts of the specific infrastructures of electricity (with electric light as proxy), tap water, and telecommunications (with possession of phone as proxy) in rural China. Due to data unavailability, transportation could not be included. Unlike the majority of previous studies that focus on efficiency effect and also rely on highly aggregated data, this paper will employ disaggregated data. Disaggregation here means using observations at the individual level and using specific physical indicators of infrastructure. The problem associated with aggregated

2 Data in this paragraph are all from National Bureau of Statistics of China (accessed March 7, 2015).

data at the country, province, or community level is well known and does not require elaboration. Arguments for using specific physical indicators rather than monetary indicators of infrastructure can be found in Straub (2008). Finally, actual access or use of these specific infrastructures is separately modeled, although they are expressed as binary variables. This avoids the difficulty in distinguishing between the issues of availability, accessibility, and affordability.

Another contribution of this paper is to point out the deficiency of the conventional approach to modeling inequality, including those estimating the Kuznets curve or testing the Kuznets hypothesis. Consequently, as discussed later in the paper, the few studies that did explore the distributive effects of infrastructure may be misleading. Alternative analytical approaches are called for and one of them is proposed in Section 3 of this paper.

Applying the proposed simple but appropriate approach to a large set of panel data compiled from China Health and Nutrition Survey database (CHNS), we find that rural telephone, tap water, and lighting/electricity infrastructures help improve per capita rural income in general. This general impact must have helped contain China's inequality rise because the rural–urban gap constitutes a dominant component of the overall income distribution (Wan 2007). More importantly, the relatively poor are found to benefit more, demonstrating that rural infrastructure can improve within-rural income inequality. Further, males, the more educated, and the more experienced gain more than their counterparts from all three rural infrastructures, particularly from telephones. Finally, some of these subpopulation effects became more significant in recent years and are larger in inland China.

The structure of the paper is as follows. Literature review is provided in the following section. This is followed by proposing a simple analytical framework for estimating the distributive effects of infrastructure in Section 3, where we also briefly show the deficiency of the conventional approach to inequality modeling. Section 4 presents empirical results and discussions. Section 5 conducts robustness checks. Finally, Section 6 concludes.

2. Literature Review

The economic literature on infrastructure began with research efforts to explain the positive correlation between the development of infrastructure, such as railroads, and rapid economic growth in the early days of industrial economies, including Western Europe, Japan, and the United States (Banerjee et al. 2012). More recently, there is an increasing recognition that infrastructure plays an important role in promoting growth and poverty reduction in less developed countries (Gramlich 1994).

The majority of existing studies have focused on the efficiency or growth effect of public expenditure on infrastructure (Gramlich 1994). For example, the pioneering work of Aschauer (1989) concluded that non-military public capital stock, particularly in terms of transport and water infrastructure, is more important than military spending in explaining productivity change in post-war US. Barro (1990) was among the first to introduce public expenditure into economic growth models and argued that public expenditure, represented by infrastructure, can induce endogenous growth.

Economists also explored various transmission mechanisms from infrastructure to growth. Firstly, Aschauer (1989), Easterly and Rebelo (1993) and Canning and Pedroni (2004) focused on the productivity channel, finding long-run positive productivity effects of public investment on infrastructure. Clarke et al. (2015) concluded that firm growth and productivity are substantially higher when Internet access is greater and when firms use the Internet more intensively. Secondly, infrastructure is found to help promote internal and external trade, thus enlarging firm's market size. This trade channel was examined and confirmed by Bougheas et al. (1999) for the European Union (EU), Duranton et al. (2014) for U.S., Faber (2014) for China, Cosar and Demir (2015) for Turkey, and Donaldson (2015) for India, implying that infrastructure can help reduce trade costs. Also see the survey of Xu (2011). Thirdly, infrastructure may facilitate migration. Using Geographic Information System (GIS) data, Atack et al. (2010) discovered that infrastructure investment in the 18th century helped promote urbanization in West and Central America. Lu et al. (2016) found that access to telecommunications, especially landline phones, is positively correlated with the likelihood of rural-to-urban migration in China. Urbanization in turn brought about economic growth. Finally, a possible channel is that development of infrastructure might enhance market competition (Du, Wei, and Xie 2013).

These growth effects were confirmed for specific infrastructures, including telecommunications in the US and other industrial countries (Cronin et al. 1991; Röller and Waverman 2001); highways, water supply, and sewerage in the US (Morrison and Schwartz 1996); and transportation in Organisation for Economic Co-operation and Development (OECD) and non-OECD economies (Canning 1999; Demetriades and Mamuneas 2000). For developing countries, apart from analytical studies by Binswanger et al. (1993) and Hulten et al. (2006) on India, and Banerjee et al. (2012) on China, a consensus seems to have emerged in the business, policy, and even academic communities that the slower growth in India compared to China can be attributed to India's poorer infrastructure. For a general discussion on the detrimental growth effects of poor infrastructure in developing countries, see Moccero (2008).

Another strand of literature directly estimates the poverty impacts of infrastructure. Gibson and Rozelle (2003) found that in rural Papua New Guinea, regions further away from major roads had more severe poverty. Dercon (2005) argued that lack of access to infrastructure prevented households from moving out of poverty. However, these poverty impacts can be easily deduced from the positive growth impacts of infrastructure, holding inequality constant. Thus, poverty reduction alone does not necessarily mean improvement in income distribution. For example, remarkable growth and poverty reduction have occurred in China since the late 1970s, but inequality emerged as a major socioeconomic problem in the 1990s, becoming even more serious over time.

In contrast to the sizable and growing literature on the growth effect of infrastructure, research outputs on its distributive or inequality effect are scarce and problematic (see Section 3). To our knowledge, only four studies examined such distributive impacts. Relying on cross-country regressions, Calderón and Chong (2004) and Calderón and Serven (2014) found that income inequality (as measured by the Gini coefficient) was negatively associated with increased availability and quality of infrastructure. Banerjee et al. (2012) explored the relationship between transportation infrastructure and the income Gini at the county level in China. Contrary to Calderón and Chong (2004) and Calderón and Serven (2014), they discovered that access to infrastructure caused rises in inequality. Finally, Zhang and Xu (2016) found that girls benefited more than boys from China's water treatment program which helped eliminate the gender gap in education in treated villages. This study provides indirect evidence of infrastructure's distributive impacts. As the following section shows, existing attempts to directly model inequality, including those testing the Kuznets hypothesis, are likely to suffer from serious specification and estimation biases.

3. Analytical Framework and Data

3.1 Analytical Framework

One of the conventional methods to estimate the growth or efficiency impacts of infrastructure is through production function modeling, where infrastructure is included in addition to the usual input variables such as capital and labor. This is clearly inapplicable when individual data are used either because of the unavailability of capital observations or due to the fact that labor input is difficult to measure at the individual level. A natural alternative is to utilize the Mincer earnings function, augmented with infrastructure variables. Let *i* index individuals and *t* index years, our baseline model can be specified as:

$$y_{it} = Ln(Inc_{it}) = \gamma_0 + \gamma_1 Sch_{it} + \gamma_2 Exp_{it} + \gamma_3 Inf_{it} + \gamma_4 X'_{it} + \phi_i + \varphi_t + u_{it} \tag{3.1}$$

where Inc denotes personal disposable income, *Sch* denotes years of schooling, and *Exp* = *Max (0, Age – Sch –* 7) denotes experience, Inf denotes infrastructure, *X* denotes control variables, ϕ denotes individual fixed effect which partly helps solve omitted variable bias. In detailed, the distribution of infrastructure access would be random across space, indicating that locational characteristics are an important part of the equation. For example, differences in the local (village) access to a certain infrastructure has to do with its distance to a nearby city, as being close to a large city could mean better access to education, markets, and job opportunities, all of which lead to higher wage incomes. The individual fixed effect incorporates the impact of the distance to a nearby city on income. φ denotes year fixed effect and u denotes the usual random error term. In this paper, infrastructure variables take *0-1* values only (1 with infrastructure and 0 without). Therefore, γ_3 measures the impact of infrastructure on personal disposable income for the treatment group.

Model (3.1) can be used to estimate the impact of infrastructure on income in general. To analyze the question of who gains more from which infrastructure, interactive variables can be included:

$$y_{it} = \alpha_0 + \alpha_1 Inf_{it} + \alpha_2 y_{i,t-1} + \alpha_3 y_{i,t-1} \times Inf_{it} + Controls + v_{it} \quad (3.2)$$

where v_{it} is the sum of fixed effects and the random error term. The definition of *Controls* is self-explanatory. When $y_{i,t-1} \times Inf_{it}$ is added, its coefficient captures the income impact of infrastructure on the relatively poor *vs* the relatively rich. To illustrate, infrastructure's impact on income in (3.2) can be expressed as:

$$E_{Inf} = E(y_{it}|Inf_{it} = 1) - E(y_{it}|Inf_{it} = 0) = \alpha_1 + \alpha_3 y_{i,t-1} \quad (3.3)$$

Obviously, the impact E_{Inf} contains two parts: (1) α_1, which is the general impact of infrastructure on income; and (2) $\alpha_3 y_{i,t-1}$, which measures the distributive impact of infrastructure on income: if $\alpha_3 > 0$, individuals with higher income in the past (higher $y_{i,t-1}$) gain more from infrastructure, indicating worsening income distribution caused by infrastructure; if $\alpha_3 < 0$, individuals with lower income in the past (lower $y_{i,t-1}$) gains more from infrastructure; $\alpha_3 = 0$ indicates absence of the distributive impact of infrastructure. By the same token, $Sch \times Inf$ can be added to explore if the better-educated gain more or less. In this paper, we will also consider gender, working experience and marriage effects of infrastructure.[3]

Our approach outlined above differs from the direct modeling of inequality adopted by Calderón and Chong (2004), Calderón and Serven (2014) and Banerjee et al. (2012). In their papers, the inequality indicator of Gini coefficient (*Gini*) is regressed on infrastructure with coefficient α_1 plus control variables X with coefficients Γ. That is:

$$Gini = \alpha_0 + \alpha_1 E(Inf) + X'\Gamma + v, \quad (3.4)$$

[3] We have also tried another specification by tabulating individuals' last period wages into 10 deciles and interacting the 10 indicators for these deciles with the infrastructure variables. The results remain robust.

where $E(Inf)$ represents the average infrastructure of a country or county. However, model (3.4) is likely to produce misleading results. To illustrate this, we simplify (3.1) as:

$$Y = Inc = \beta_0 + \beta_1 Inf + Z'\Delta + u, \tag{3.5}$$

where Z denotes all K control variables with coefficients $\Delta = \{\delta_k\}$. Then the "true" *Gini* index of Inc can be derived as (see Wan 2004):

$$Gini(Y) = \sum_k \delta_k E(Z_k)/E(Y)\, Con(Z_k) + \beta_1\, E(Inf)/E(Y)\, Con(Inf) \tag{3.6}$$

where E is the expectation operator, *Con* denotes the concentration coefficient, which can be computed using:

$$Con(Z_k) = -2Cov\left(Z_k/E(Z_k), \left(1 - F(Y)\right)\right), \tag{3.7}$$

where F denotes the cumulative distribution function of Y. Note that $Con(Z_k)$ does not change when $E(Z_k)$ changes.

The "true" marginal impact of $E(Inf)$ on the *Gini* can be easily derived from (3.6):

$$\frac{\partial Gini(Y)}{\partial E(Inf)} = \beta_1 [E(Y)]^{-1} [Con(Inf) - Gini(Y)] \tag{3.8}$$

Clearly, the "true" distributive effect given by (3.8) differs from α_1 which represents the estimated distributive effect of infrastructure under the conventional approach used by Calderon and Serven (2014) and others.

Although our proposed approach does not generate direct estimates of the impact of infrastructure on inequality indicators like the *Gini*, it does provide insights regarding which subpopulation groups gain more from infrastructure. To properly identify and directly estimate the infrastructure impacts on an inequality index, the inequality accounting framework of Wan (2004) can be considered.

3.2 Data

To empirically estimate model (3.2), data are compiled from the database of CHNS, a longitudinal household survey conducted by the Carolina Population Center of the University of North Carolina and the National Institute of Food Safety of the Chinese Center for Disease Control and Prevention. It is carried out by relevant city/county anti-epidemic stations under the provincial Food Inspection Services in China. Although the CHNS is designed to collect information on health, nutrition, and family planning, it does contain detailed income and infrastructure data.

The survey covers the years 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, and 2011. For each year, approximately 4,400 households in nine provinces were surveyed, involving interviews with some 16,000 individuals. The provinces are Liaoning, Heilongjiang, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi, and Guizhou, mostly in Eastern or East-Central China. Households from Heilongjiang were added in 1997, and those from Liaoning were not surveyed in 1997. A multistage, random clustering design is adopted to draw samples. Within each province, counties are stratified by income (low, medium, and high) and a weighted sampling scheme is used to select four counties in each province.

In this paper, we focus on individuals who were over 18 years old in the survey year and resided in rural areas. We do not exclude the old-aged who usually work as long as their health permits. This may lead to mortality selection bias, which will be dealt with later in the paper when conducting the robustness check. Income observations are deflated or inflated with 2009 as the base year. Infrastructure is measured as a binary variable in terms of actual use or consumption:

1. Whether he/she uses telephone(s)

2. Whether he/she consumes tap water

3. Whether he/she uses (electric) lights

Table 1 provides definitions of major variables and Table 2 tabulates their summary statistics. In CHNS, personal income is not a simple division of household income by household size. It represents actual earnings of individuals, with subsidies, gifts, rent and in-kind payments excluded. Referring to the last two rows of Panel A of Table 2, the sample sizes are quite large, no less than 270,000. Limiting to rural observations results in 48,024 observations. When matched with other variables, the number of observations used for model estimations will be smaller. One may notice the maximum value of 94.67 for the experience variable. This is rather large and will be dealt with later in the paper when conducting the robustness check.

[Insert Table 1 approximately here]

The bottom half of Table 2 presents the summary statistics of the infrastructure variables. Access to electric lights is high, increasing from almost 90% in 1989 to 99 percent in 2011. Such a high access rate implies that its impacts might be hard to identify and estimate, particularly in the later years. Access to tap water reached 81 percent in 2011, from a low level of 36% in 1989. The least accessible or consumed infrastructure is telephone. The survey did not cover this variable until 1997, presumably due to negligible use of private phones in rural China before then. Also, there seem to be some problems with this variable as the rate of access reached 62 percent in 2004 and then declined. This data issue will be handled later in the paper by dropping those individuals who lost phone access.

[Insert Table 2 approximately here]

4. Empirical Results and Discussions

4.1 The Baseline Model

The baseline model of (3.1) is estimated using OLS with fixed effects and the standard errors clustered at the household level. Table 3 reports the estimation results. Broadly speaking, most of the estimated parameters have the expected signs, although the schooling variable is not significant in any of the equations. Also, lights are negatively correlated with income, but the parameters are not significantly different from 0. The insignificance does not necessarily mean that electricity is unimportant. As mentioned earlier, the electricity/lighting impact is hard to identify or estimate given the rather small variation in the observations. In any case, having access to tap water and telephone can significantly increase the income of rural laborers, averaging about 3–4 percentage points.

Because the rural–urban gap accounts for 70% or more of the total inequality in China (Wan 2007) and the rural average income is one-third or less of the urban counterpart, the general positive income effects imply that infrastructure investment in rural China had played a role in helping contain the worsening income distribution in China. In other words, without these investments, China's alarming inequality could be worse. Nevertheless, this finding cannot be used to answer the key question of who gained more from infrastructural development, an issue to be addressed in Section 4.2.

It is interesting to note that the return to working experience is estimated to be around 7 percent when evaluated at its sample mean of 20.4 years (see Table 2). This is higher than the estimated return to schooling. This larger-than-expected estimate may be caused by the omitted variable bias in model (3.1). For example, returns to schooling and experience may be conditional on infrastructure, which requires inclusion of interactive variables in the model. This problem will be dealt with in the next subsection.

[Insert Table 3 approximately here]

4.2 Distributive Impacts for Population Sub-groups

Next, we use model (3.2) to estimate the distributive impact of infrastructure. There are two potential problems with model (3.2). One is omitted relevant variable bias. Individuals could earn more because infrastructure enhanced the earning power of those with better education or more work experience. Omission of this interactive effect may cause bias in estimation. One way to alleviate this problem is to control as many individual-specific variables as possible, such as schooling, experience, gender, marriage, and their interactions with infrastructure.

Another problem is the possible presence of endogeneity since (3.2) is a dynamic panel model with fixed effects. Therefore, we employ system GMM of Blundell and Bond (1998) to estimate the model as the sample is large enough to overcome the problems of unstable and potentially biased estimators as recent studies have raised (Roodman, 2007, 2009a, 2009b; Bazzi and Clemens, 2013; Bun and Windmeijer, 2010). The lagged income variable and its interaction with infrastructure are set to be endogenous (GMM-style) variables. The lag length is chosen when second-order autocorrelation disappears. It is also important to ensure that the model is not over identified, which can be formally tested by the Hansen statistics.

Panel A of Table 4 reports the estimation results. Contrasting with Table 3, the models fit the data better in terms of both signs of parameter estimates and levels of significance, even for the light or electricity equation. However, the light or electricity model remains less reliable. For example, the estimated coefficient of AR(1) in the electricity model exceeds unity. In any case, the estimated rate of return to schooling is 4-5%, systematically lower than the counterparts in Table 3, indicating possible biases caused by endogeneity in model (3.1). More importantly, as the α_3 parameter is significantly negative in all equations, it can be inferred again that the relatively poorer gained more from all the three basic infrastructures. One implication is that infrastructure as a public good can be used as a policy tool for combating inequality, not just for promoting growth.

[Insert Table 4 approximately here]

According to (3.3), the log-income difference between those with and without infrastructure is simply $\alpha_1 + \alpha_3 y_{i,t-1}$. These differences are evaluated at the average log-income of those without the infrastructure so the computed differences exactly represent the impacts generated by the presence of infrastructure. Panel B in Table 4 reports these impacts[4]. Clearly, they differ across infrastructure. Telephone contributes slightly more to income growth than tap water (e.g. 37.84% > 33.76%). This is not surprising as telephones provide market and employment information to rural household, opening up more opportunities. On the other hand, the impact of tap water is more indirect and gradual, largely reflected in improvement in the health status.

Now, we consider add additional interactive variables to gauge possible differences in the

[4] Given that the light/electricity model is less unreliable, this computation is only done for telephone and tap water.

impacts of infrastructure for population sub-groups classified by schooling, gender, marital status and experience. Taking schooling as an example, the model to be estimated can be written as

$$y_{it} = \gamma_0 + \gamma_1 Sch_{it} \times Inf_{it} + Controls \tag{4.1}$$

In the above model, $\gamma_1 > 0$ means the better-educated benefit more from infrastructure, and *vice versa*.

The same identification and estimation strategy adopted to obtain Table 4 is repeated here. Panel A of Table 5 reports the estimation results. The coefficients for the interactive variables indicate that the better-educated benefit more than the less educated, so do the male than female and the more experienced than less experienced. Premium for the married is positive but insignificant. These are all consistent with a priori expectations. In particular, the earlier finding that telephone exerted larger impacts is still valid (30.58% > 22.62%, see bottom panel of Table 5).

[Insert Table 5 approximately here]

Contrary to the result in Table 4, the coefficient estimate of the interactive term between schooling and infrastructure is larger than that of the experience–infrastructure term. Also, unlike in Table 4, the coefficients of schooling and experience as reported in Table 5 are no longer significant. One possible reason is that without basic infrastructure, residents in rural China have few opportunities to engage in income earning activities. In other words, returns to education and experience are likely to be conditional on the presence of basic infrastructure, as confirmed by the significant coefficients of the interaction terms of school and experience with infrastructure. That is, infrastructure helps open up more opportunities for those with better education or more experience. Thus, more investment in education in rural China is called for in order to fully explore the synergy between human capital and infrastructure.

5. Further Discussion and Robustness Check

In this section, we use subsample data to estimate models for different time periods and different areas. Robustness check is then carried out by redefining the experience variable, by alleviating mortality selection bias, by reconstructing the telephone data, and by addressing possible reverse causality. Hereafter, we only consider models for telephone and tap water as data for electricity or lights contain limited variations.

5.1 Subsample Results

Recall that trade is one of the transmission channels from infrastructure to growth and distribution. It is thus useful to separately model the infrastructure impacts for pre-WTO (1989-2000) and post-WTO (2000-2011) periods. Since observations on telephone were not collected until 1997 it is not feasible to estimate the telephone model by GMM for the period of 1989-2000. Instead, data for the period of 1989-2006 will be used.

Table 6 presents the estimation results. It is clear that the signs of parameter estimates are all consistent with those in Table 5 with only a couple of exceptions. But these few exceptional cases are associated with insignificant coefficients. More importantly, both the general income effects and distributive effects became stronger over time as far as telephone is concerned, judged by the increases in the absolute values of parameter estimates. The exactly opposite occurred to tap water. These results may partly reflect the fact that tap water accessibility is always higher than that of telephone except in the crisis year of 2009.

[Insert Table 6 approximately here]

To examine the robustness to different areas, data for east (Liaoning, Jiangsu, and Shandong), central (Heilongjiang, Henan, Hubei, and Hunan) and west (Guangxi and Guizhou) China are used to estimate the telephone and tap water equations. As Table 7 shows, the estimates are broadly consistent with those in Tables 5 and 6, reinforcing the earlier robustness check results. Also, it is interesting to note that both the general income and distributive effects are larger in inland China than in east China, likely due to the already higher accessibility rates in coastal China.

[Insert Table 7 approximately here]

5.2 Robustness Check for Measurement Errors

Referring to Table 2, the maximum value for the experience variable reached 94.7 years, which seems to be problematic. To examine robustness to potential errors in this variable, we cap its values at certain ages. That is, we redefine this variable by assuming that experience does not change anymore after an individual reaches 65, 70 or 80 respectively. The results remain robust according to Table 8 which confirms that the better educated, the more experienced, and the male gain more from rural infrastructure.

[Insert Table 8 approximately here]

Next, our data may suffer from mortality selection bias, as the old individual in the sample may possess characteristics that are different from those who have passed away (Fitzgerald, Gottschalk and Moffitt 1998). To examine sensitivity to this problem, we drop those observations when the age variable is greater than 65, 70 or 80. Observations for individuals before they reached these ages are kept in the sample. Again, the results remain robust according to Table 9.

[Insert Table 9 approximately here]

The third possible measurement error relates to the observed decline in telephone accessibility after 2004. This could be caused by possible substitution of mobiles for landline phones. In the early years, mobile phones were rare in rural China and the later surveys may have failed to fully account for the substitution. One imperfect but acceptable way to deal with this problem is to drop observations for those who previously had access to a phone but “lost” access later. Observations before he/she lost the access are retained. The results are reported in Table 10. Once again, the results are robust. Note that the mortality selection bias is corrected for models in columns 2-4 of Table 10.

[Insert Table 10 approximately here]

5.3 Endogeneity (Reverse Causality)

One remaining problem is possible endogeneity as consumption of infrastructure may depend on affordability or income. To address this problem, we construct an infrastructure variable at the village level by replacing the binary observations on infrastructure by the average access rates at the village level. This new infrastructure variable becomes continuous in the interval [0, 1]. This is justified as infrastructure has spillover effect. A household without direct access to tap water may still benefit from tap water if the access rate of the corresponding village is high enough. More importantly, any single individual cannot significantly influence the access rate of the whole village. We will also cluster the standard error at the village level. In this way, reversed causality or endogeneity can be alleviated. Table 11 reports the estimation results. Once again, our results are robust.[5]

[Insert Table 11 approximately here]

6. Summary and Conclusion

This paper is motivated by the shortage of research on the distributive effect of infrastructure, despite an increasing literature on its growth impacts since late 1980s. In addition, the need to focus on the distributive impact arises from the universal and growing discontent with rising inequality, coupled with huge spending on infrastructure in many countries.

Our paper began with a short discussion on the deficiency of the conventional approach to inequality modeling. We then propose a simple but useful framework for estimating both growth and distributive effects. When applied to data from rural China, it is found that all infrastructures (telephone, tap water and to some extent electricity/lights) helped raise rural income and the growth effects became larger in later years. More importantly, income gains differ for different population groups. By and large, the relatively poor, male, more experienced, better educated and to a less extent the married share more of the gains, relative to their

[5] It's noted that the coefficients of $y_{i,t-1} \times Inf_{it}$ become smaller. Possible explanation is smaller variation of infrastructures in the sample when aggregated into village level.

counterparts. The telephone effects are stronger than tap water and the infrastructure impacts are more significant in inland China than elsewhere. It is useful to point out that the tap water effect is likely to be indirect, with long lags. Such long-run effects may not be fully captured by our models. Proper identification and reliable estimation of such long run effects also require time series data that span more years than the current CHNS database can provide.

References

Aschauer, David Alan. "Is public expenditure productive?" *Journal of Monetary Economics* 23, no. 2 (1989): 177-200.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A. Margo. "Did railroads induce or follow economic growth." *Social Science History* 34, no. 2 (2010): 171-97.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. "On the road: Access to transportation infrastructure and economic growth in China." No. w17897. National Bureau of Economic Research, 2012.

Barro, Robert J. "Government spending in a simple model of endogeneous growth." *Journal of Political Economy* (1990): 103-S125.

Binswanger, Hans P., Shahidur R. Khandker, and Mark R. Rosenzweig. "How infrastructure and financial institutions affect agricultural output and investment in India." *Journal of Development Economics* 41, no. 2 (1993): 337-366.

Blundell, Richard, and Stephen Bond. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87, no. 1 (1998): 115-143.

Bougheas, Spiros, Panicos O. Demetriades, and Edgar LW Morgenroth. "Infrastructure, transport costs and trade." *Journal of International Economics* 47, no. 1 (1999): 169-189.

Calderón, César, and Alberto Chong. "Volume and quality of infrastructure and the distribution of income: an empirical investigation." *Review of Income and Wealth* 50, no. 1 (2004): 87-106.

Calderón, César, and Luis Servén. "The Effects of infrastructure development on growth and income distribution." *Annals of Economics and Finance* 15, no.2 (2014): 521-534.

Canning, David, and P. Pedroni. "The contribution of infrastructure to aggregate output." *The World Bank Policy Research Working Paper* 2246 (1999).

Canning, David, and Peter Pedroni. "The effect of infrastructure on long run economic growth." *Harvard University* (2004): 1-30.

Clarke, George, Christine Z. Qiang, L. Colin Xu. "The internet as a general-purpose technology." *Economics Letters* 135 (2015): 24-27.

Cosar, A. K. and B. Demir. "Domestic road infrastructure and international trade: Evidence from Turkey." *Journal of Development Economics* 118 (2016): 232-244.

Cronin, Francis J., Edwin B. Parker, Elisabeth K. Colleran, and Mark A. Gold. "Telecommunications infrastructure and economic growth: An analysis of causality." *Telecommunications Policy* 15 no. 6 (1991): 529-535.

Demetriades, Panicos O., and Theofanis P. Mamuneas. "Intertemporal output and employment effects of public infrastructure capital: evidence from 12 OECD economies." *Economic Journal* 110 no. 465 (2000): 687-712.

Dercon, Stefan. "Risk, poverty and vulnerability in Africa." *Journal of African Economies* 14 no. 4 (2005): 483-488.

Donaldson, D. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*, forthcoming, 2015.

Du, Qingyuan, Shang-Jin Wei, and Peichu Xie. Roads and the Real Exchange Rate. No. w19291. National Bureau of Economic Research, 2013.

Duranton, G., P. M. Morrow and M. A. Turner. "Roads and Trade: Evidence from the US." *Review of Economic Studies* 81 no. 2 (2014): 681-724.

Easterly, William, and Sergio Rebelo. "Fiscal policy and economic growth." *Journal of Monetary Economics* 32 no. 3 (1993): 417-458.

Faber, B. "Trade integration, market size, and industrialization: Evidence from China’s national trunk highway system." *Review of Economic Studies* 81 no.3 (2014): 1046-1070.

Fitzgerald, J., P. Gottschalk, and R. Moffitt. "An analysis of sample attrition in panel data." *Journal of Human Resources* 33 (1998): 251–99.

Gibson, John, and Scott Rozelle. "Poverty and access to roads in Papua New Guinea." *Economic Development and Cultural Change* 52 no. 1 (2003): 159-185.

Gramlich, Edward M. "Infrastructure investment: A review essay." *Journal of Economic Literature* (1994): 1176-1196.

Hulten, Charles R., Esra Bennathan, and Sylaja Srinivasan. "Infrastructure, externalities, and economic development: a study of the Indian manufacturing industry." *World Bank Economic Review* 20 no. 2 (2006): 291-308.

Lu, Yi, Huihua, Xie, and Lixin. C. Xu. "Telecommunication externality on migration: evidence from Chinese villages." *China Economic Review* 39 (2016): 77-90.

Moccero, Diego. Improving the Business and Investment Climate in Indonesia. No. 638. OECD Publishing, 2008.

Morrison, Catherine J., and Amy Ellen Schwartz. "Public infrastructure, private input demand, and economic performance in New England manufacturing." *Journal of Business & Economic Statistics* 14 no. 1 (1996): 91-101.

National Bureau of Statistics of China, China Statistical Yearbook. (China Statistics Press, Various year)

Nerlove, M. "Experimental evidence on the estimation of dynamic economic relations from a time series of cross sections." *Economic Studies Quarterly* 18 (1967): 42-74.

Piketty, Thomas. "Capital in the 21st Century." *Cambridge: Harvard Uni* (2014).

Röller, Lars-Hendrik, and Leonard Waverman. "Telecommunications infrastructure and economic development: A simultaneous approach." *American Economic Review* (2001): 909-923.

Shorrocks, Anthony, and Guanghua Wan. "Spatial decomposition of inequality." *Journal of Economic Geography* 5 no. 1 (2005): 59-81.

Straub, Stéphane. *Infrastructure and Growth in Developing Countries*. Vol. 4460. World Bank Publications, 2008.

Wan, Guanghua. "Accounting for income inequality in rural China: a regression-based approach." *Journal of Comparative Economics* 32 no. 2 (2004): 348-363.

Wan, Guanghua. "Understanding regional poverty and inequality trends in China: methodological issues and empirical findings." *Review of Income and Wealth* 53 no. 1 (2007): 25-34.

Wan, Guanghua, ed. *Inequality and growth in modern China*. Oxford University Press, 2008a.

Wan, Guanghua, ed. *Understanding inequality and poverty in China: Methods and applications*. Palgrave Macmillan, 2008b.

Wang, Chen, Guanghua Wan, and Dan Yang. "Income inequality in the People's Republic of China: Trends, determinants, and proposed remedies." *Journal of Economic Surveys* 28 no. 4 (2014): 686-708.

World Bank, *World Development Report*. Washington, DC, the World Bank, 1994.

Xu, Lixin Colin. "The effects of business environments on development: A survey of new firm-level evidence." *World Bank Research Observer* 26 no. 2 (2011): 310-340.

Zhang, Jing, and Lixin, C. Xu. "The long-run effects of treated water on education: The rural drinking water program in China." *Journal of Development Economics* 122(2016): 1-15.

Table 1. Variable Definition

Variable	Definition
Ln(Inc)	Log value of total individual income inflated to 2009
Telephone	Dummy=1 with telephone
Tap Water	Dummy=1 with tap water
Light	Dummy=1 with electric lights
Sch	Years of schooling
Exp	Years of experience, estimated as *max*(0, age- School-7)
Exp^2	The squared value of experience
Gender	Dummy=1 if the individual is male
Marry	Dummy=1 if the individual is married
East	Dummy=1 if the individual resides in eastern China
Mid	Dummy=1 if the individual resides in central China

Table 2. Summary Statistics

Panel A. Data Description					
Variable	N	Mean	Std. Dev.	Min	Max
Ln(Inc)	48024	8.396	1.271	0.271	13.434
Ln(Inc_{t-1})	38846	8.237	1.224	0.271	13.434
Telephone	66554	0.490	0.500	0	1
Tap water	90169	0.623	0.485	0	1
Light	90062	0.979	0.142	0	1
Sch	60176	6.330	4.073	0	18
Exp	92657	20.434	21.006	0	94.67
Gender	62833	0.490	0.500	0	1
Marry	65461	0.757	0.429	0	1
East	92657	0.311	0.463	0	1
Mid	92657	0.412	0.492	0	1

Panel B. Infrastructure Improvement									
Variables	1989	1991	1993	1997	2000	2004	2006	2009	2011
Telephone=1	/	/	/	1575	3516	6094	7287	6083	7367
Telephone=0	/	/	/	6407	5588	3663	4930	6761	7283
Telephone Accessibility (%)	/	/	/	19.73	38.62	62.46	59.65	47.36	50.29
Tap Water=1	2827	3620	3867	4534	5556	6213	8371	9427	11792
Tap Water=0	4936	4571	3801	3486	3581	3564	3831	3342	2850
Tap Water Accessibility (%)	36.42	44.19	50.43	56.53	60.81	63.55	68.60	73.83	80.54
Light=1	6933	7719	7530	7952	9006	9744	12140	12706	14481
Light=0	820	476	132	67	87	25	46	37	161
Light Accessibility (%)	89.42	94.19	98.28	99.16	99.04	99.74	99.62	99.71	98.90

Source: Authors estimation based on CHNS data.

Table 3. Baseline Estimation Results: The General Effect of Infrastructure on Income

Ln(Inc)	Telephone		Tap water		Light	
Inf	0.0365*	0.0377*	0.0425**	0.0427**	-0.00154	-0.00387
	(0.0215)	(0.0215)	(0.0197)	(0.0198)	(0.0483)	(0.0483)
Sch	0.0684	0.0652	0.0751	0.0726	0.0780	0.0756
	(0.0457)	(0.0434)	(0.0502)	(0.0484)	(0.0524)	(0.0506)
Exp	0.0817*	0.0788*	0.100**	0.0967**	0.103**	0.1000**
	(0.0452)	(0.0429)	(0.0499)	(0.0480)	(0.0521)	(0.0503)
Exp^2	-0.000540***	-0.000545***	-0.000749***	-0.000733***	-0.000751***	-0.000736***
	(6.12e-05)	(6.34e-05)	(4.32e-05)	(4.50e-05)	(4.32e-05)	(4.50e-05)
Controls	No	Yes	No	Yes	No	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	30,090	29,885	45,757	45,500	45,718	45,460
R-squared	0.167	0.168	0.183	0.184	0.183	0.184

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered at the household level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4. The Distributive Impacts of Infrastructure

Panel A: Regression

$Ln(Inc_t)$	Telephone		Tap water		Light	
$Ln(Inc_{t-1})$	0.642***	0.538**	0.695*	0.734*	6.765*	5.603*
	(0.226)	(0.227)	(0.401)	(0.408)	(3.975)	(2.892)
$Ln(Inc_{t-1})$*Inf	-0.644***	-0.559***	-0.679*	-0.725*	-6.798*	-5.649**
	(0.205)	(0.204)	(0.369)	(0.375)	(3.957)	(2.879)
Inf	5.670***	4.941***	5.695*	6.081**	53.67*	44.62**
	(1.721)	(1.721)	(2.990)	(3.040)	(31.10)	(22.62)
Sch	0.0497***	0.0446***	0.0492***	0.0411***	0.0666***	0.0576***
	(0.00690)	(0.00634)	(0.00844)	(0.00706)	(0.00375)	(0.00352)
Exp	0.000935	0.00322	0.0199***	0.0178***	0.0247***	0.0231***
	(0.00576)	(0.00556)	(0.00661)	(0.00585)	(0.00494)	(0.00393)
Exp^2	-5.57e-05	-0.000109	-0.000339***	-0.000324***	-0.000378***	-0.000378***
	(8.61e-05)	(8.47e-05)	(9.41e-05)	(8.72e-05)	(7.79e-05)	(6.20e-05)
Controls	No	Yes	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)-p	0.132	0.220	0.120	0.112	0.105	0.067
Hansen-p	0.109	0.125	0.900	0.903	0.220	0.504
N	20,259	20,121	29,346	29,174	29,325	29,152

Panel B: Infrastructure's Effect on Income

Infrastructure	Mean $Ln(Inc_{t-1})$ without the infrastructure	Income Effect
Telephone	8.162	37.84%
Tap Water	7.922	33.76%

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered at household level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$;

4) $Income\ Effect = \theta_2 Ln(Inc_{it-1}) + \theta_3$

Table 5. The Distributive Impacts of Infrastructure by population sub-groups

Panel A: Regression

$Ln(Inc_t)$	Telephone		Tap water		Light	
$Ln(Inc_{t-1})$	0.695***	0.615***	0.721*	0.775*	7.445*	6.720**
	(0.224)	(0.231)	(0.414)	(0.425)	(4.227)	(3.254)
$Ln(Inc_{t-1})$*Inf	-0.714***	-0.652***	-0.723*	-0.786**	-7.482*	-6.768**
	(0.208)	(0.213)	(0.390)	(0.399)	(4.212)	(3.243)
Sch*Inf	0.0650***	0.0523***	0.0774**	0.0753**	0.576*	0.548*
	(0.0166)	(0.0162)	(0.0313)	(0.0297)	(0.333)	(0.285)
Exp*Inf	0.00600***	0.00448**	0.0104***	0.00958***	0.0510	0.0499
	(0.00213)	(0.00211)	(0.00370)	(0.00346)	(0.0486)	(0.0441)
Gender*Inf		0.153***		0.0838**		-0.463
		(0.0382)		(0.0400)		(0.762)
Marry*Inf		0.0632		0.171		2.039
		(0.0730)		(0.136)		(1.330)
Inf	5.620***	5.082***	5.250*	5.610**	54.49*	47.59**
	(1.608)	(1.610)	(2.865)	(2.840)	(30.72)	(22.87)
Sch	0.0146	0.0156	-0.000450	-0.00765	-0.501	-0.483*
	(0.0148)	(0.0140)	(0.0282)	(0.0256)	(0.330)	(0.282)
Exp	-0.00380	-0.00125	0.0122	0.0102	-0.0276	-0.0284
	(0.00646)	(0.00637)	(0.00933)	(0.00840)	(0.0508)	(0.0454)
Exp^2	-3.70e-05	-8.18e-05	-0.000324***	-0.000303***	-0.000350***	-0.000345***
	(8.55e-05)	(8.61e-05)	(0.000103)	(9.69e-05)	(8.95e-05)	(7.15e-05)
Controls	No	Yes	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.000	0.000	0.001	0.001	0.000	0.000
AR(2)-p	0.111	0.164	0.125	0.111	0.080	0.068
Hansen-p	0.189	0.224	0.801	0.833	0.252	0.592
N	20,259	20,121	29,346	29,174	29,325	29,152

Panel B: Infrastructure Effect on Income

Infrastructure	Mean $Ln(Inc_{t-1})$ without the infrastructure	Income Effect
Telephone	8.162	30.58%
Tap Water	7.922	22.62%

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered in household level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$;

4) $Income\ Effect = \theta_2 + \theta_3 Ln(Inc_{it-1}) + \theta_4 Mean(Sch) + \theta_5 Mean(Exp) + \theta_6 Mean(Gender) + \theta_7 Mean(Marry)$.

Table 6. Robustness Check: Different Time Periods

$Ln(Inc_t)$	Telephone		Tap water		
	1989-2006	2000-2011	1989-2000	1989-2006	2000-2011
$Ln(Inc_{t-1})$	0.993***	1.176**	0.736	0.538	0.567*
	(0.310)	(0.532)	(0.574)	(0.383)	(0.334)
$Ln(Inc_{t-1})$*Inf	-1.049***	-1.180**	-0.767	-0.574^{+}	-0.596*
	(0.300)	(0.493)	(0.549)	(0.366)	(0.317)
Sch*Inf	0.0520***	0.0867**	0.0482*	0.0499**	0.0592***
	(0.0164)	(0.0354)	(0.0277)	(0.0218)	(0.0206)
Exp*Inf	0.00154	0.00656**	0.00588**	0.00506**	0.00415*
	(0.00250)	(0.00324)	(0.00261)	(0.00208)	(0.00213)
Gender*Inf	0.284***	0.204***	0.0971**	0.0746*	0.0868
	(0.0536)	(0.0685)	(0.0472)	(0.0428)	(0.0632)
Marry*Inf	0.158	0.183	0.219	0.152	0.109
	(0.100)	(0.115)	(0.196)	(0.126)	(0.0973)
Inf	8.207***	9.129**	5.562	4.145	4.508*
	(2.275)	(3.767)	(3.950)	(2.634)	(2.426)
Sch	0.0110	-0.0134	0.00685	0.0160	0.0136
	(0.0158)	(0.0316)	(0.0261)	(0.0190)	(0.0185)
Exp	-0.00220	-0.0144	0.0241*	0.0239***	0.00320
	(0.00884)	(0.0109)	(0.0126)	(0.00762)	(0.00470)
Exp^2	-5.26e-05	0.000102	-0.000489***	-0.000466***	-0.000164**
	(0.000122)	(0.000150)	(0.000174)	(0.000101)	(6.45e-05)
Controls	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.000	0.001	0.017	0.001	0.000
AR(2)-p	0.349	0.128	0.189	0.195	0.333
Hansen-p	0.315	0.466	0.514	0.357	0.210
N	13,741	16,735	16,313	22,802	16,742

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered in household level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$;

4) $^{+}$ $p < 0.12$.

Table 7. Robustness Check: Different Areas

$Ln(Inc_t)$	Telephone			Tap water		
	East	Middle	West	East	Middle	West
$Ln(Inc_{t-1})$	1.119^{+}	0.290*	0.610**	0.819	0.691*	0.996
	(0.704)	(0.174)	(0.295)	(0.591)	(0.364)	(0.673)
$Ln(Inc_{t-1})$*Inf	-1.118*	-0.357**	-0.697**	-0.784	-0.716**	-1.031^{+}
	(0.666)	(0.161)	(0.276)	(0.557)	(0.340)	(0.643)
Sch*Inf	0.0819*	0.0347**	0.0586**	0.0693*	0.0742***	0.0973*
	(0.0446)	(0.0147)	(0.0235)	(0.0402)	(0.0254)	(0.0553)
Exp*Inf	0.000105	0.00408	0.00602*	0.00851	0.0107***	0.0101**
	(0.00392)	(0.00313)	(0.00336)	(0.00521)	(0.00414)	(0.00486)
Gender*Inf	0.124	0.141**	0.0574	0.184	0.140**	-0.148
	(0.130)	(0.0591)	(0.0772)	(0.134)	(0.0549)	(0.151)
Marry*Inf	0.224	0.0244	-0.000332	0.252	0.144	0.155
	(0.223)	(0.0956)	(0.108)	(0.224)	(0.130)	(0.234)
Inf	9.044*	2.735**	5.450**	5.612	5.021**	7.483
	(5.185)	(1.193)	(2.121)	(3.928)	(2.414)	(4.640)
Sch	-0.0213	0.0348***	0.0187	-0.0101	-0.00275	-0.0268
	(0.0402)	(0.0117)	(0.0200)	(0.0347)	(0.0202)	(0.0515)
Exp	-0.00683	0.0129	0.000423	0.00123	0.0140	0.0134
	(0.00636)	(0.00786)	(0.00993)	(0.00805)	(0.0105)	(0.0121)
Exp^2	3.44e-05	-0.000281***	-9.70e-05	-0.000173**	-0.000372***	-0.000346**
	(8.69e-05)	(0.000107)	(0.000134)	(8.67e-05)	(0.000131)	(0.000137)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.002	0.000	0.000	0.005	0.001	0.003
AR(2)-p	0.270	0.358	0.234	0.106	0.273	0.281
Hansen-p	0.196	0.110	0.753	0.420	0.600	0.215
N	6,529	8,089	5,503	9,947	11,256	7,971

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status;
2) Robust standard errors in parentheses and are clustered in household level;
3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$;
4) $^{+}$ $p < 0.12$.

Table 8. Robustness Check: Redefining the Experience Variable

Ln(Inc_t)	Telephone			Tap water		
	Experience capped at			Experience capped at		
	Age = 65	Age =70	Age = 80	Age = 65	Age = 70	Age = 80
Ln(Inc_{t-1})	0.555**	0.573**	0.608***	0.772*	0.794*	0.782*
	(0.238)	(0.235)	(0.231)	(0.421)	(0.422)	(0.424)
Ln(Inc_{t-1})*Inf	-0.597***	-0.613***	-0.646***	-0.784**	-0.804**	-0.792**
	(0.219)	(0.217)	(0.214)	(0.396)	(0.396)	(0.398)
Sch*Inf	0.0476***	0.0492***	0.0518***	0.0749**	0.0760**	0.0754**
	(0.0174)	(0.0167)	(0.0162)	(0.0308)	(0.0302)	(0.0297)
Exp*Inf	0.00383	0.00405*	0.00442**	0.00958**	0.00952**	0.00945***
	(0.00252)	(0.00226)	(0.00211)	(0.00415)	(0.00375)	(0.00346)
Gender*Inf	0.151***	0.151***	0.153***	0.0813**	0.0840**	0.0845**
	(0.0376)	(0.0379)	(0.0382)	(0.0395)	(0.0398)	(0.0400)
Marry*Inf	0.0504	0.0540	0.0614	0.171	0.178	0.174
	(0.0708)	(0.0718)	(0.0728)	(0.130)	(0.133)	(0.136)
Inf	4.690***	4.805***	5.039***	5.607**	5.755**	5.665**
	(1.637)	(1.632)	(1.612)	(2.789)	(2.808)	(2.834)
Sch	0.0166	0.0162	0.0155	-0.00966	-0.0101	-0.00815
	(0.0141)	(0.0140)	(0.0140)	(0.0256)	(0.0255)	(0.0256)
Exp	0.00663	0.00394	-0.000544	0.0169	0.0141	0.0106
	(0.00884)	(0.00779)	(0.00653)	(0.0110)	(0.00975)	(0.00857)
Exp^2	-0.000218*	-0.000167	-9.41e-05	-0.000432***	-0.000375***	-0.000309***
	(0.000128)	(0.000109)	(8.88e-05)	(0.000137)	(0.000118)	(9.96e-05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.000	0.000	0.000	0.001	0.001	0.001
AR(2)-p	0.247	0.218	0.171	0.110	0.101	0.107
Hansen-p	0.232	0.229	0.224	0.821	0.828	0.834
N	20,121	20,121	20,121	29,174	29,174	29,174

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered in household level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$;

4) $^{+}$ $p < 0.12$.

Table 9. Robustness Check: Drop Observations for the Old

Ln(Inc_t)	Telephone			Tap water		
	Observation dropped when			Observation dropped when		
	age > 65	age > 70	age > 80	age > 65	age > 70	age > 80
Ln(Inc_{t-1})	0.222	0.387^{+}	0.550**	0.297	0.448**	0.544
	(0.241)	(0.243)	(0.237)	(0.222)	(0.227)	(0.364)
Ln(Inc_{t-1})*Inf	-0.298	-0.452**	-0.595***	-0.354*	-0.493**	-0.573*
	(0.223)	(0.225)	(0.219)	(0.210)	(0.214)	(0.342)
Sch*Inf	0.0262	0.0377**	0.0474***	0.0424**	0.0526***	0.0591**
	(0.0183)	(0.0179)	(0.0166)	(0.0172)	(0.0172)	(0.0261)
Exp*Inf	-0.00153	0.00237	0.00390*	0.00598*	0.00713**	0.00763**
	(0.00349)	(0.00278)	(0.00219)	(0.00317)	(0.00278)	(0.00329)
Gender*Inf	0.136***	0.128***	0.150***	0.0479	0.0519*	0.0695*
	(0.0364)	(0.0364)	(0.0378)	(0.0313)	(0.0309)	(0.0356)
Marry*Inf	0.0216	0.0319	0.0512	0.0864	0.110	0.108
	(0.0730)	(0.0748)	(0.0727)	(0.0806)	(0.0820)	(0.116)
Inf	2.512	3.615**	4.670***	2.536*	3.531**	4.115*
	(1.634)	(1.671)	(1.652)	(1.462)	(1.505)	(2.424)
Sch	0.0334**	0.0242	0.0181	0.0180	0.0101	0.00582
	(0.0152)	(0.0150)	(0.0143)	(0.0141)	(0.0142)	(0.0222)
Exp	0.0155	0.0123	0.00403	0.0256***	0.0238***	0.0174**
	(0.00949)	(0.00851)	(0.00705)	(0.00675)	(0.00601)	(0.00783)
Exp^2	-0.000318**	-0.000299**	-0.000165*	-0.000527***	-0.000508***	-0.000398***
	(0.000129)	(0.000117)	(9.61e-05)	(8.42e-05)	(7.29e-05)	(9.08e-05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.000	0.000	0.000	0.000	0.000	0.000
AR(2)-p	0.755	0.626	0.264	0.457	0.128	0.221
Hansen-p	0.228	0.174	0.237	0.138	0.119	0.350
N	17,892	19,025	19,979	26,553	27,927	29,009

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered in household level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$;

4) $^{+}$ $p < 0.12$.

Table 10. Robustness Check: Telephone

$Ln(Inc_t)$		Observations dropped when			
	Full Sample	Age > 65	Age > 70	Age > 80	Full Sample
$Ln(Inc_{t-1})$	0.389*	0.0708	0.248	0.311	0.411*
	(0.216)	(0.256)	(0.233)	(0.223)	(0.221)
$Ln(Inc_{t-1})$*Inf	-0.452**	-0.162	-0.326	-0.383*	-0.476**
	(0.207)	(0.246)	(0.224)	(0.214)	(0.212)
Sch*Inf	0.0453***	0.0199	0.0283**	0.0295***	0.0344***
	(0.0111)	(0.0143)	(0.0119)	(0.0111)	(0.0110)
Exp*Inf	0.00499***	-0.00193	0.00149	0.00260	0.00351**
	(0.00169)	(0.00304)	(0.00186)	(0.00176)	(0.00171)
Gender*Inf		0.208***	0.207***	0.216***	0.222***
		(0.0375)	(0.0363)	(0.0363)	(0.0364)
Marry*Inf		-0.0246	-0.0259	-0.0226	-0.00841
		(0.0790)	(0.0783)	(0.0777)	(0.0783)
Inf	3.569**	1.492	2.677	3.092*	3.781**
	(1.631)	(1.826)	(1.700)	(1.634)	(1.623)
Sch	0.0269***	0.0409***	0.0358***	0.0362***	0.0336***
	(0.00972)	(0.0121)	(0.0102)	(0.00958)	(0.00950)
Exp	0.00650	0.0219**	0.0148**	0.0124**	0.00692
	(0.00555)	(0.00932)	(0.00691)	(0.00612)	(0.00562)
Exp^2	-0.000199**	-0.000412***	-0.000314***	-0.000278***	-0.000192**
	(8.07e-05)	(0.000131)	(0.000102)	(9.03e-05)	(8.27e-05)
Controls	No	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.000	0.000	0.000	0.000	0.000
AR(2)-p	0.348	0.777	0.719	0.557	0.330
Hansen-p	0.140	0.168	0.118	0.153	0.142
N	17,311	15,466	16,930	17,189	17,311

Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered in household level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11. Robustness Check: Endogeneity

$Ln(Inc_t)$	Telephone		Tap Water		Light	
$Ln(Inc_{t-1})$	0.338**	0.306*	0.511*	0.445^{+}	0.583**	0.502*
	(0.156)	(0.157)	(0.272)	(0.276)	(0.271)	(0.275)
$Ln(Inc_{t-1})$*Inf	-1.71e-05**	-1.61e-05**	-2.15e-05*	-1.93e-05*	-2.75e-05**	-2.41e-05*
	(7.78e-06)	(7.63e-06)	(1.11e-05)	(1.09e-05)	(1.28e-05)	(1.26e-05)
Sch*Inf	0.0487***	0.0418***	0.0314***	0.0267***	0.0404	-0.00172
	(0.0113)	(0.0121)	(0.00787)	(0.00881)	(0.0354)	(0.0370)
Exp*Inf	0.00549*	0.00425	0.00537***	0.00450**	-0.00610	-0.0128
	(0.00307)	(0.00324)	(0.00172)	(0.00179)	(0.00984)	(0.0112)
Gender*Inf		0.181**		0.0603		0.571***
		(0.0743)		(0.0489)		(0.163)
Marry*Inf		-0.174*		-0.0313		0.0108
		(0.101)		(0.0873)		(0.259)
Inf	0.0791	0.206	-0.0616	-0.00241	0.231	0.327
	(0.185)	(0.205)	(0.0876)	(0.113)	(0.452)	(0.425)
Sch	0.0224**	0.0195**	0.0184	0.0171	0.00271	0.0409
	(0.00901)	(0.00921)	(0.0141)	(0.0126)	(0.0359)	(0.0372)
Exp	0.00430	0.00549	0.00893	0.00998	0.0174	0.0252**
	(0.00549)	(0.00556)	(0.00974)	(0.00893)	(0.0117)	(0.0119)
Exp^2	-0.000169**	-0.000197***	-0.000239*	-0.000260*	-0.000214*	-0.000243**
	(7.53e-05)	(7.64e-05)	(0.000143)	(0.000134)	(0.000124)	(0.000117)
Controls	No	Yes	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)-p	0.000	0.000	0.001	0.001	0.000	0.000
AR(2)-p	0.170	0.203	0.075	0.127	0.043	0.087
Hansen-p	0.127	0.179	0.311	0.240	0.556	0.458
N	20,379	20,239	29,455	29,280	29,455	29,280

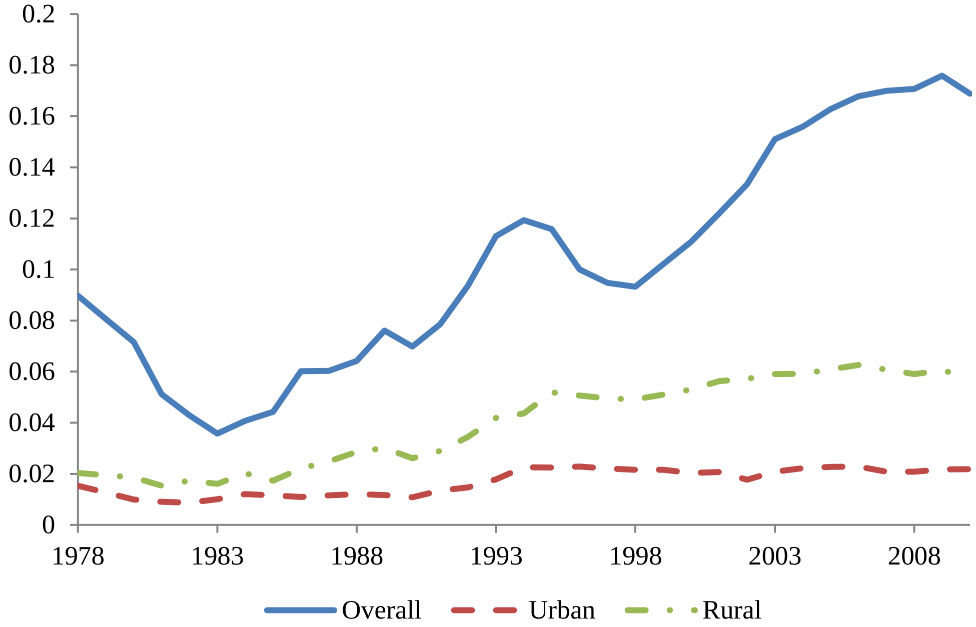
Note: 1) The controlled variables include years of schooling, years of experience, gender, marriage status, dummy of eastern region, and dummy of middle region;

2) Robust standard errors in parentheses and are clustered in village level;

3) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$;

4) $^{+}$ $p < 0.12$.

Figure 1. Regional Inequality in China: Theil Index



Source: Estimated based on group income data from National Bureau of Statistics of China (various years).