

Wan, Guanghua; Zhang, Xun

**Working Paper**

## Who gains more from which infrastructure in rural People's Republic of China?

ADB Working Paper, No. 540

**Provided in Cooperation with:**

Asian Development Bank Institute (ADBI), Tokyo

*Suggested Citation:* Wan, Guanghua; Zhang, Xun (2015) : Who gains more from which infrastructure in rural People's Republic of China?, ADB Working Paper, No. 540, Asian Development Bank Institute (ADBI), Tokyo

This Version is available at:

<https://hdl.handle.net/10419/115357>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**ADB Working Paper Series**

**Who Gains More from Which Infrastructure  
in Rural People's Republic of China?**

---

Guanghua Wan  
and Xun Zhang

No. 540  
July 2015

**Asian Development Bank Institute**

Guanghua Wan is Director of Research at the Asian Development Bank Institute.

Xun Zhang is an assistant professor at Beijing Normal University.

The authors are grateful to the Natural Science Foundation of China for its support through NSFC project 71133004. Comments from participants of various seminars, GDN, and ADBI workshops are acknowledged.

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of ADBI, ADB, its Board of Directors, or the governments they represent. ADBI does not guarantee the accuracy of the data included in this paper and accepts no responsibility for any consequences of their use. Terminology used may not necessarily be consistent with ADB official terms.

Working papers are subject to formal revision and correction before they are finalized and considered published.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. ADBI encourages readers to post their comments on the main page for each working paper (given in the citation below). Some working papers may develop into other forms of publication.

Suggested citation:

Wan, G., and X. Zhang. 2015. Who Gains More from Which Infrastructure in Rural People's Republic of China? ADBI Working Paper 540. Tokyo: Asian Development Bank Institute. Available: <http://www.adb.org/publications/who-gains-more-which-infrastructure-rural-prc/>

Please contact the authors for information about this paper.

Email: [gwan@adbi.org](mailto:gwan@adbi.org), [zhangxun@bnu.edu.cn](mailto:zhangxun@bnu.edu.cn)

Asian Development Bank Institute  
Kasumigaseki Building 8F  
3-2-5 Kasumigaseki, Chiyoda-ku  
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500

Fax: +81-3-3593-5571

URL: [www.adbi.org](http://www.adbi.org)

E-mail: [info@adbi.org](mailto:info@adbi.org)

© 2015 Asian Development Bank Institute

**Abstract**

The importance of infrastructure in economic development has been increasingly recognized by governments, development institutions, and the research community. Despite a sizable literature on its efficiency and growth effects, the distributive impacts of infrastructure have been largely overlooked, with a few recent exceptions. This is regrettable, particularly given the overwhelming concern about inequality and inclusive growth all over the world. This paper will: (i) demonstrate the deficiency of conventional approaches to modelling inequality; (ii) extend the Mincer earnings function so that both growth and distributive effects of infrastructure can be evaluated; and (iii) fit the extended model to a large sample of individual-level data from rural People's Republic of China (PRC) over the period of 1989–2011, providing estimates of growth and the 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 the PRC'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 telephones. Finally, some of these subpopulation effects have become more significant in recent years and are larger in central PRC, possibly because infrastructure helps open up more opportunities for those with better education or more experience. 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.

**JEL Classification:** D31, H52, H54

## Contents

1.	Introduction.....	3
2.	A Brief Literature Review .....	5
3.	Analytical Framework and Data .....	6
4.	Empirical Results and Discussions .....	10
4.1	Baseline Model .....	10
4.2	Distributive Impacts for Population Subgroups.....	11
5.	Further Discussion Robustness Checks .....	14
5.1	Subsample Results .....	14
5.2	Robustness Check for Measurement Errors.....	17
5.3	Endogeneity (Reverse Causality).....	19
6.	Conclusions .....	20
	References .....	22

## 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 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, 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 fast-growing economies spend a significant proportion of government budget on infrastructure. Even industrialized countries are confronted with infrastructure upgrading.<sup>1</sup> 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.

The People's Republic of China (PRC) offers a natural setting to shed light on the distributive impacts of infrastructure. The PRC'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 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 the PRC 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 in rural PRC.<sup>2</sup>

Meanwhile, worsening income distribution has been ranked among the top three most serious socioeconomic and policy issues for decades in the PRC (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 PRC increased faster than its urban counterpart. The much steeper slope of the overall inequality curve than either of the

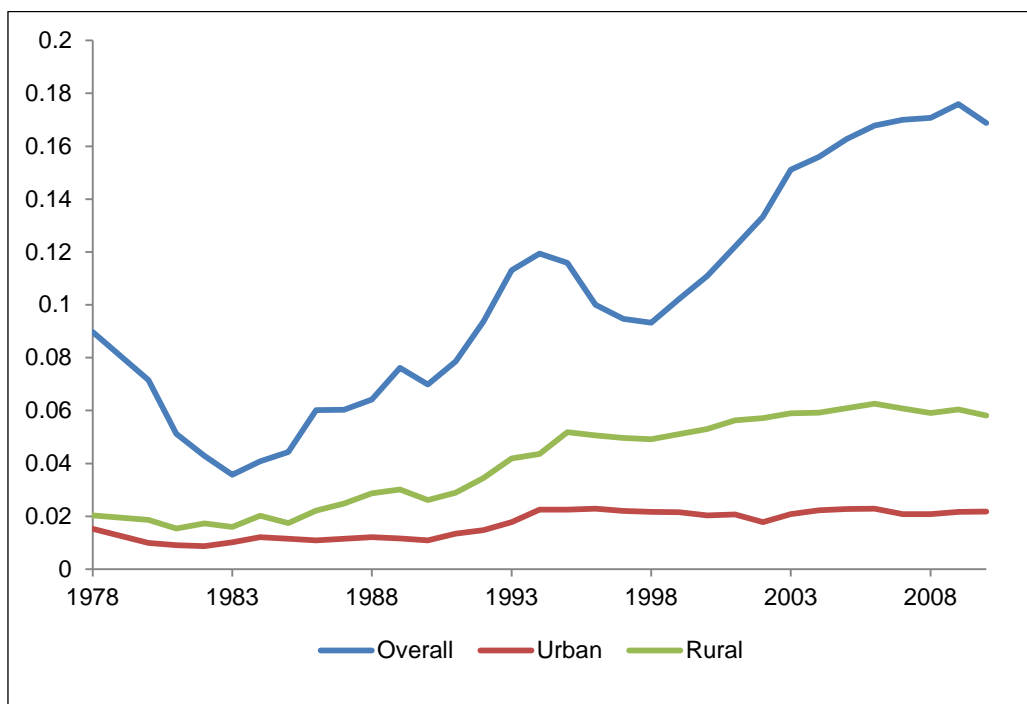
---

<sup>1</sup> 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.

<sup>2</sup> Data in this paragraph is from the National Bureau of Statistics of China (accessed 7 March 2015).

other two curves implies large and growing urban–rural gaps. This contrasts with the fact that the PRC’s massive infrastructure investment aims at narrowing down regional as well as urban–rural gaps. It is thus important to analytically explore the distributive impacts of infrastructure.

**Figure 1: Regional Inequality in the People’s Republic of China: Theil Index**



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

This paper represents one of the first efforts to analyze the distributive impacts of the specific infrastructures of electricity (with electric lights as a proxy), tap water, and telecommunications (with possession of a phone as a proxy) in rural PRC. Due to data unavailability, transportation could not be included. Unlike the majority of previous studies that focus on efficiency effects 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 data at the country, province, or community level is well known and does not require elaboration. Preference for using specific physical indicators rather than monetary indicators of infrastructure is discussed by 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 the 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 the

PRC'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. This equalizing impact became more significant in later years. Further, males, the more educated, and the more experienced gain more than their counterparts from all three rural infrastructures, particularly from telephones. Finally, the distributive impacts are more significant in central PRC.

The structure of the paper is as follows. A brief literature review is provided in the following section. This is followed by proposing a simple analytical framework to gauge 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. A BRIEF 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, Duflo, and Qian 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 effects of public expenditure on infrastructure (Gramlich 1994). For example, the pioneering work of Aschauer (1989) concluded that non-military public capital stock, particularly 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.

Like Aschauer (1989), Easterly and Rebelo (1993) and, more recently, Canning and Pedroni (2004) have focused on productivity impacts, finding evidence of long-run positive effects of public investment in infrastructure. Using data from European Union (EU) countries, Bougheas, Demetriades, and Morgenroth (1999) found a positive correlation between the level of infrastructure and the volume of trade, implying that infrastructure helps reduce trade costs. 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. Urbanization in turn brought about economic growth. Through theoretical and empirical analyses, Du, Wei, and Xie (2013) demonstrated the enhancement effect of infrastructure on competition and price, and therefore on the real exchange rate.

These productivity or economic 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).

Turning to developing countries, apart from analytical studies by Binswanger, Khandker, and Rosenzweig (1993) and Hulten, Bennathan, and Srinivasan (2006) on India, and Banerjee, Duflo, and Qian (2012) on the PRC, a consensus seems to have



emerged in the business, policy, and even academic communities that the slower growth in India compared to the PRC can be largely attributed to its poorer infrastructure. For a general discussion on the detrimental growth effects of poor infrastructure in developing countries, see Moccerro (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 prevents 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 the PRC 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 effects of infrastructure, research output on the distributive or inequality effects are scarce and problematic (see Section 3). To our knowledge, only three studies have examined such distributive impacts. Relying on cross-country regressions, Calderón and Chong (2004) and Calderón and Servén (2014) found that income inequality (as measured by the Gini coefficient) was negatively associated with increased availability and quality of infrastructure. In another paper, Banerjee et al. (2012) explored the relationship between transportation infrastructure and the income Gini at the county level in the PRC. Contrary to Calderón and Chong (2004) and Calderón and Servén (2014), they discovered that access to infrastructure caused rises in inequality. 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

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 inappropriate when individual data is used either because of unavailability of capital 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, and our baseline model can be specified as:

$$y_{it} = \ln(Inc_{it}) = \gamma_0 + \gamma_1 Sch_{it} + \gamma_2 Exp_{it} + \gamma_3 Inf_{it} + \gamma_4 X'_{it} + \phi_i + \varphi_t + u_{it} \quad (1)$$

where  $Inc$  denotes personal disposable income,  $Sch$  denotes years of schooling,  $Exp = \max(0, Age - Sch - 7)$  denotes experience,  $Inf$  denotes infrastructure,  $X$  denotes control variables,  $\phi$  denotes individual fixed effect,  $\varphi$  denotes year effect and  $u$  denotes the error term. In this paper, infrastructure variables take a value of 0 or 1, only (1 with infrastructure and 0 without).

$\gamma_3$  measures the impact of infrastructure on personal disposable income for the treatment group. To illustrate, we define  $\omega_{it}$  as:

$$\omega_{it} \triangleq y_{it} - \gamma_1 Sch_{it} - \gamma_2 Exp_{it} - \gamma_4 X'_{it} \quad (2)$$

Clearly,  $\omega_{it}$  represents the personal disposable income determined by infrastructure, the unobservable provincial fixed effect, and the time fixed effect, with the effects of other variables purged.

Now, consider two persons, A and B. In period 1, both are without infrastructure ( $Inf = 0$ ). In period 2, B is equipped with infrastructure ( $Inf = 1$ ), while A is not ( $Inf = 0$ ). Then, the differences over time for A and B are respectively:

$$D_A = E(\omega_{A2}) - E(\omega_{A1}) = \varphi_2 - \varphi_1 \quad (3)$$

$$D_B = E(\omega_{B2}) - E(\omega_{B1}) = \gamma_3 + \varphi_2 - \varphi_1 \quad (4)$$

It is clear that  $\gamma_3$  is equivalent to the difference-in-difference estimator of infrastructure's effect on income—that is,  $D_B - D_A$ .

Model (1) can be used to assess the impact of infrastructure on income in general. To explore the topical question of who gains more from which infrastructure, interactive variables can be included. When  $Ln(Inc_{it-1}) \times Inf$  is added, its coefficient captures the income impact of infrastructure on the relatively poor versus the relatively rich. A negative coefficient indicates that those who were poorer previously, gain more. By the same token,  $Sch \times Inf$  can be added to explore whether the better-educated gain more or less. In this paper, we also consider the gender, working experience, and marriage effects of infrastructure. It is noted that when an interactive variable such as  $Ln(Inc_{it-1}) \times Inf$  is added, it is customary to have the individual variables—for instance,  $Ln(Inc_{it-1})$  and  $Inf$ —included in the model as well. In particular, the inclusion of the lagged income variable helps accommodate serial autocorrelation, typically inherent in the income variable. Its coefficient signifies convergence (inequality reducing) or divergence (inequality increasing). This differs from the conventional model of convergence or catch-up where income in the initial year rather than the previous year is used. Inclusion of the initial income in the model would make fixed effects estimation impossible.

Our approach differs from the direct modeling of inequality adopted by Calderón and Chong (2004), Calderón and Servén (2014) and Banerjee, Duflo, and Qian (2012). In their papers, the inequality indicator of the Gini coefficient ( $Gini$ ) is regressed on infrastructure with coefficient  $\alpha_1$  plus control variables  $X$  with coefficients  $\Gamma$ —that is:

$$Gini = \alpha_0 + \alpha_1 E(Inf) + X'\Gamma + v, \quad (5)$$

where  $E(Inf)$  represents the average infrastructure of a country or county. However, model (5) is likely to produce misleading results. To illustrate this, we simplify (1) as:

$$Y = Inc = \beta_0 + \beta_1 Inf + Z'\Delta + u, \quad (6)$$

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]):

$$Gini(Y) = \sum_k \delta_k E(Z_k) / E(Y) Con(Z_k) + \beta_1 E(Inf) / E(Y) Con(Inf) \quad (7)$$

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

$$Con(Z_k) = -2Cov\left(Z_k / E(Z_k), (1 - F(Y))\right), \quad (8)$$

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 (7):

$$\text{Marginal distributive impact} = [\beta_1 E(Y) - \beta_1^2 E(\text{Inf})] \text{Con}(\text{Inf}) / E^2(Y) \quad (9)$$

It is beyond the scope of this paper to provide the proof, but the “true” distributive effect given by (9) clearly differs from  $\alpha_1$ , which represents the estimated distributive effect of infrastructure under the conventional approach used by Calderón and Servén (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.

To empirically estimate model (1) with additional interactive variables, data are compiled from the 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. Although the CHNS is designed to collect information on health, nutrition, and family planning, it does contain detailed income and infrastructure data that can be used to address the topical question.

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 PRC. Households from Heilongjiang have been added since 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 did 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 a telephone(s)
2. Whether he/she consumes tap water
3. Whether he/she can use (electric) lights

Table 1 provides definitions of major variables and Table 2 tabulates their summary statistics. Referring to the last two rows of Panel A of Table 2, the sample sizes are quite large, no less than 270,000. We limit our study to rural areas, resulting in only 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.

**Table 1: Variable Definitions**

Variable	Definition
Ln(Inc)	Log value of personal income
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 $\max(0, \text{age} - \text{School} - 7)$
Exp <sup>2</sup>	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 PRC
Mid	Dummy=1 if the individual resides in central PRC

PRC = People's Republic of China.

Source: Authors' compilation based on CHNS.

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 presence of private phones in rural PRC 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.

**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 <sub>t-1</sub> )	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: Author's estimation based on CHNS data.

## 4. EMPIRICAL RESULTS AND DISCUSSIONS

### 4.1 Baseline Model

The baseline model of (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 the PRC (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 PRC had played a role in helping contain the worsening income distribution in the PRC. In other words, without these investments, the PRC's alarming inequality could be worse. Nevertheless, this finding cannot be used to answer the topical question of who gained more from infrastructural development, an issue to be addressed in Section 4.2.

It is interesting to note that a 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. The larger-than-expected estimate may be caused by the omitted variable bias in model (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.

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

	Telephone		Tap water		Light	
Inf	0.0365* (0.0215)	0.0377* (0.0215)	0.0425** (0.0197)	0.0427** (0.0198)	-0.00154 (0.0483)	-0.00387 (0.0483)
Sch	0.0684 (0.0457)	0.0652 (0.0434)	0.0751 (0.0502)	0.0726 (0.0484)	0.0780 (0.0524)	0.0756 (0.0506)
Exp	0.0817* (0.0452)	0.0788* (0.0429)	0.100** (0.0499)	0.0967** (0.0480)	0.103** (0.0521)	0.1000** (0.0503)
Exp <sup>2</sup>	-0.000540*** (6.12e-05)	-0.000545*** (6.34e-05)	-0.000749*** (4.32e-05)	-0.000733*** (4.50e-05)	-0.000751*** (4.32e-05)	-0.000736*** (4.50e-05)
Control variables	No	Yes	No	Yes	No	Yes
Individual effect	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	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) Robust standard errors in parentheses, clustered at the household level.

2) \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Source: Authors' estimations based on CHNS data.

## 4.2 Distributive Impacts for Population Subgroups

As discussed earlier, the model to be used to assess the convergence or divergence impact of infrastructure can be expressed as (disturbance term suppressed):

$$\ln(Inc_{it}) = \theta_0 + \theta_1 \ln(Inc_{it-1}) + \theta_2 \ln(Inc_{it-1}) \times Inf + \theta_3 Inf + Controls \quad (10)$$

where  $\theta_2 > 0$  implies that the relatively rich benefit more from infrastructure, while  $\theta_2 < 0$  represents the opposite.

There are two issues with the model in equation (10). One is omitted relevant variable bias. People become richer probably because infrastructure enhanced the potential earning power of those with better education or more work experience. This will result in 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 presence of endogeneity since (10) is a dynamic panel model with fixed effects. Therefore, we employ system GMM of Blundell and Bond (1998) to estimate (10). 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 ( $p$ -value  $> 0.10$ ).

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 level 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 equation (1). More importantly, as the  $\theta_2$  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.

**Table 4: Distributive Impacts of Infrastructure**

Panel A: Regression						
	Telephone		Tap water		Light	
Ln(Inc <sub>t-1</sub> )	0.642*** (0.226)	0.538** (0.227)	0.695* (0.401)	0.734* (0.408)	6.765* (3.975)	5.603* (2.892)
Ln(Inc <sub>t-1</sub> )*Inf	-0.644*** (0.205)	-0.559*** (0.204)	-0.679* (0.369)	-0.725* (0.375)	-6.798* (3.957)	-5.649** (2.879)
Inf	5.670*** (1.721)	4.941*** (1.721)	5.695* (2.990)	6.081** (3.040)	53.67* (31.10)	44.62** (22.62)
Sch	0.0497*** (0.00690)	0.0446*** (0.00634)	0.0492*** (0.00844)	0.0411*** (0.00706)	0.0666*** (0.00375)	0.0576*** (0.00352)
Exp	0.000935 (0.00576)	0.00322 (0.00556)	0.0199*** (0.00661)	0.0178*** (0.00585)	0.0247*** (0.00494)	0.0231*** (0.00393)
Exp <sup>2</sup>	-5.57e-05 (8.61e-05)	-0.000109 (8.47e-05)	-0.000339*** (9.41e-05)	-0.000324*** (8.72e-05)	-0.000378*** (7.79e-05)	-0.000378*** (6.20e-05)
Control variables	No	Yes	No	Yes	No	Yes
Time effect	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
Observations	20,259	20,121	29,346	29,174	29,325	29,152
Panel B: Infrastructure's Effect on Income						
Infrastructure	Mean Ln(Inc <sub>t-1</sub> ) without the infrastructure		Income effect			
Telephone	8.162		37.84%			
Tap Water	7.922		33.76%			

Note: 1) Robust standard errors in parentheses, clustered at household level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

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

Source: Author's estimation based on CHNS data.

According to equation (10), the log-income difference between those with and without infrastructure is simply  $\theta_2 Ln(Inc_{it-1}) + \theta_3$ . 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 the impacts.<sup>3</sup> Clearly, they differ across infrastructure. Telephone contributes slightly more to income growth than tap water (for instance, 37.84% > 33.76%). This is not surprising as telephones provide market and employment information to rural households, opening up more opportunities. On the other hand, the impact of tap water is more indirect and gradual, largely reflected in the health status.

Now, let us add additional interactive variables into the model, examining the possible differences in the impacts of infrastructure for population subgroups classified by schooling, gender, marital status, and experience. Taking schooling as an example, the model to be estimated can be written as

$$Ln(Inc_{it}) = \gamma_0 + \gamma_1 Sch \times Inf + Controls \tag{11}$$

In model (11),  $\gamma_1 > 0$  means the better-educated benefit more from infrastructure

<sup>3</sup> Given that the light/electricity model is less unreliable, this evaluation will only be done for telephone and tap water.

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, as do males than females, and the more experienced than the less experienced. The premium for the married is positive but insignificant. These are all consistent with *a priori* expectations. In particular, the earlier finding that telephones exerted larger impacts is still valid (30.58% > 22.62%, see bottom panel of Table 5).

**Table 5: Distributive Impacts of Infrastructure by Population Subgroup**

Panel A: Regression						
	Telephone		Tap water		Light	
Ln(Inc <sub>t-1</sub> )	0.695*** (0.224)	0.615*** (0.231)	0.721* (0.414)	0.775* (0.425)	7.445* (4.227)	6.720** (3.254)
Ln(Inc <sub>t-1</sub> )*Inf	-0.714*** (0.208)	-0.652*** (0.213)	-0.723* (0.390)	-0.786** (0.399)	-7.482* (4.212)	-6.768** (3.243)
Sch*Inf	0.0650*** (0.0166)	0.0523*** (0.0162)	0.0774** (0.0313)	0.0753** (0.0297)	0.576* (0.333)	0.548* (0.285)
Exp*Inf	0.00600*** (0.00213)	0.00448** (0.00211)	0.0104*** (0.00370)	0.00958*** (0.00346)	0.0510 (0.0486)	0.0499 (0.0441)
Gender*Inf		0.153*** (0.0382)		0.0838** (0.0400)		-0.463 (0.762)
Marry*Inf		0.0632 (0.0730)		0.171 (0.136)		2.039 (1.330)
Inf	5.620*** (1.608)	5.082*** (1.610)	5.250* (2.865)	5.610** (2.840)	54.49* (30.72)	47.59** (22.87)
Sch	0.0146 (0.0148)	0.0156 (0.0140)	-0.000450 (0.0282)	-0.00765 (0.0256)	-0.501 (0.330)	-0.483* (0.282)
Exp	-0.00380 (0.00646)	-0.00125 (0.00637)	0.0122 (0.00933)	0.0102 (0.00840)	-0.0276 (0.0508)	-0.0284 (0.0454)
Exp <sup>2</sup>	-3.70e-05 (8.55e-05)	-8.18e-05 (8.61e-05)	-0.000324*** (0.000103)	-0.000303*** (9.69e-05)	-0.000350*** (8.95e-05)	-0.000345*** (7.15e-05)
Control variables	No	Yes	No	Yes	No	Yes
Time effect	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
Observations	20,259	20,121	29,346	29,174	29,325	29,152
Panel B: Infrastructure Effect on Income						
Infrastructure	Mean Ln(Inc <sub>t-1</sub> ) without the infrastructure			Income effect		
Telephone	8.162			30.58%		
Tap water	7.922			22.62%		

Note: 1) Robust standard errors in parentheses, clustered at household level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

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

Source: Authors' estimations based on CHNS data.

Contrary to the results 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 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 PRC 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. In other words, infrastructure helps open up more opportunities for those with better education or more experience. Thus, more investment in education in rural PRC is called for in order to fully explore the synergy between human capital and infrastructure.

## **5. FURTHER DISCUSSION ROBUSTNESS CHECKS**

In this section, we use subsample data to estimate models for different time periods and different areas. Robustness checks are 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**

It was planned to examine the robustness to timing by splitting data into two periods: 1989–2000 and 2000–2011. However, observations on telephones were not collected until the 1997 survey. Thus, for telephone, it is not feasible to estimate the model by GMM for the period of 1989–2000. Instead, data for the period 1989–2006 is used.

Table 6 presents the robustness check 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 efficiency/income effects and distributive effects became stronger over time as far as telephones are concerned, judging by the increases in the absolute values of the parameter estimates. The opposite occurred in the case of tap water. These results may partly reflect the fact that tap water accessibility is always higher than for telephones, except in the crisis year of 2009.

**Table 6: Robustness Check: Different Time Periods**

	Telephone		Tap water		
	1989–2006	2000–2011	1989–2000	1989–2006	2000–2011
Ln(Inc <sub>t-1</sub> )	0.993*** (0.310)	1.176** (0.532)	0.736 (0.574)	0.538 (0.383)	0.567* (0.334)
Ln(Inc <sub>t-1</sub> )*Inf	-1.049*** (0.300)	-1.180** (0.493)	-0.767 (0.549)	-0.574 <sup>+</sup> (0.366)	-0.596* (0.317)
Sch*Inf	0.0520*** (0.0164)	0.0867** (0.0354)	0.0482* (0.0277)	0.0499** (0.0218)	0.0592*** (0.0206)
Exp*Inf	0.00154 (0.00250)	0.00656** (0.00324)	0.00588** (0.00261)	0.00506** (0.00208)	0.00415* (0.00213)
Gender*Inf	0.284*** (0.0536)	0.204*** (0.0685)	0.0971** (0.0472)	0.0746* (0.0428)	0.0868 (0.0632)
Marry*Inf	0.158 (0.100)	0.183 (0.115)	0.219 (0.196)	0.152 (0.126)	0.109 (0.0973)
Inf	8.207*** (2.275)	9.129** (3.767)	5.562 (3.950)	4.145 (2.634)	4.508* (2.426)
Sch	0.0110 (0.0158)	-0.0134 (0.0316)	0.00685 (0.0261)	0.0160 (0.0190)	0.0136 (0.0185)
Exp	-0.00220 (0.00884)	-0.0144 (0.0109)	0.0241* (0.0126)	0.0239*** (0.00762)	0.00320 (0.00470)
Exp <sup>2</sup>	-5.26e-05 (0.000122)	0.000102 (0.000150)	-0.000489*** (0.000174)	-0.000466*** (0.000101)	-0.000164** (6.45e-05)
Control variables	Yes	Yes	Yes	Yes	Yes
Time effect	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
Observations	13,741	16,735	16,313	22,802	16,742

Note: 1) Robust standard errors in parentheses, clustered at household level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

3) <sup>+</sup> p < 0.12.

Source: Author's estimation based on CHNS data.

To examine the robustness to different areas, data for east, central, and west PRC 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 PRC than in east PRC, probably due to the already higher accessibility rate in coastal PRC.

**Table 7: Robustness Check: Different Areas**

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

Note: 1) Robust standard errors in parentheses, clustered at household level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

3) <sup>+</sup> p < 0.12.

Source: Author's estimation based on CHNS data.

### 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. 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.

**Table 8: Robustness Check: Redefining the Experience Variable**

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

Note: 1) Robust standard errors in parentheses, clustered at household level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

3) + p < 0.12.

Source: Authors' estimations based on CHNS data.

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

**Table 9: Robustness Check: Drop Observations for the Old**

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

Note: 1) Robust standard errors in parentheses, clustered at household level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

3) <sup>+</sup> p < 0.12.

Source: Authors' estimations based on CHNS data.

The third possible measurement error relates to the observed decline in telephone accessibility after 2004. This could be caused by possible confusion between landline and mobile phones. In the early years, mobile phones were rare in rural PRC, and the later surveys may have failed to fully account for the replacement of the landline by mobiles. One imperfect but acceptable way to deal with such errors 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.

**Table 10: Robustness Check: Telephone**

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

Note: 1) Robust standard errors in parentheses, clustered at household level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

Source: Authors' estimations based on CHNS data.

### 5.3 Endogeneity (Reverse Causality)

Recall that biases arising from possibly omitted variables and mortality selection have been addressed. To a certain extent, biases caused by unobservable variables are corrected via fixed effects estimations. Major measurement errors are discussed in 5.2.

One remaining problem is possible endogeneity as consumption of infrastructure may depend on affordability or income. To alleviate this reverse causality problem, we construct an infrastructure variable at the village level by replacing the binary observations on infrastructure by the average access rate to various infrastructures of relevant villages. This new infrastructure variable becomes continuous in the interval [0, 1]. This is justified as infrastructure has a 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.

**Table 11: Robustness Check: Endogeneity**

	Telephone		Tap Water		Light	
Ln(InC <sub>t-1</sub> )	0.338** (0.156)	0.306* (0.157)	0.511* (0.272)	0.445* (0.276)	0.583** (0.271)	0.502* (0.275)
Ln(InC <sub>t-1</sub> )*Inf	-1.71e-05** (7.78e-06)	-1.61e-05** (7.63e-06)	-2.15e-05* (1.11e-05)	-1.93e-05* (1.09e-05)	-2.75e-05** (1.28e-05)	-2.41e-05* (1.26e-05)
Sch*Inf	0.0487*** (0.0113)	0.0418*** (0.0121)	0.0314*** (0.00787)	0.0267*** (0.00881)	0.0404 (0.0354)	-0.00172 (0.0370)
Exp*Inf	0.00549* (0.00307)	0.00425 (0.00324)	0.00537*** (0.00172)	0.00450** (0.00179)	-0.00610 (0.00984)	-0.0128 (0.0112)
Gender*Inf		0.181** (0.0743)		0.0603 (0.0489)		0.571*** (0.163)
Marry*Inf		-0.174* (0.101)		-0.0313 (0.0873)		0.0108 (0.259)
Inf	0.0791 (0.185)	0.206 (0.205)	-0.0616 (0.0876)	-0.00241 (0.113)	0.231 (0.452)	0.327 (0.425)
Sch	0.0224** (0.00901)	0.0195** (0.00921)	0.0184 (0.0141)	0.0171 (0.0126)	0.00271 (0.0359)	0.0409 (0.0372)
Exp	0.00430 (0.00549)	0.00549 (0.00556)	0.00893 (0.00974)	0.00998 (0.00893)	0.0174 (0.0117)	0.0252** (0.0119)
Exp <sup>2</sup>	-0.000169** (7.53e-05)	-0.000197*** (7.64e-05)	-0.000239* (0.000143)	-0.000260* (0.000134)	-0.000214* (0.000124)	-0.000243** (0.000117)
Control Variables	No	Yes	No	Yes	No	Yes
Time Effect	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
Observations	20,379	20,239	29,455	29,280	29,455	29,280

Note: 1) Robust standard errors in parentheses, clustered at village level.

2) \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

3) + p < 0.12.

Source: Authors' estimations based on CHNS data.

## 6. CONCLUSIONS

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

Having outlined the deficiency of the conventional approach to inequality modeling, a simple but useful model is proposed and applied to examine the distributive effects of telephone, tap water, and to some extent electricity/lighting on individual income in rural PRC. It is found that all infrastructures helped raise rural income, with the growth effects becoming larger in later years. More importantly, income gains differ for different population groups. By and large, males, the relatively poor, the more experienced, the better educated, and to a lesser extent those who are married share more of the gains relative to their counterparts. The telephone effects are stronger than for tap water, and the infrastructure impacts are more significant in inland PRC 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, D. A. 1989. Is Public Expenditure Productive? *Journal of Monetary Economics* 23(2): 177–200.
- Atack, J., F. Bateman, M. Haines, and R. A. Margo. 2010. Did Railroads Induce or Follow Economic Growth. *Social Science History* 34, no. 2 (2010): 171–97.
- Banerjee, A., E. Duflo, and N. Qian. 2012. On the Road: Access to Transportation Infrastructure and Economic Growth in China. NBER Working Paper No. 17897. Cambridge, MA: National Bureau of Economic Research.
- Barro, R. J. 1990. Government Spending in a Simple Model of Endogeneous Growth. *Journal of Political Economy* 98(5): 103–125.
- Binswanger, H. P., S. R. Khandker, and M. R. Rosenzweig. 1993. How Infrastructure and Financial Institutions Affect Agricultural Output and Investment in India. *Journal of development Economics* 41(2): 337–366.
- Blundell, R., and S. Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87(1): 115–143.
- Bougheas, S., P. O. Demetriades, and E. L. W. Morgenroth. 1999. Infrastructure, Transport Costs and Trade. *Journal of International Economics* 47(1): 169–189.
- Calderón, C., and A. Chong. 2004. Volume and Quality of Infrastructure and the Distribution of Income: An Empirical Investigation. *Review of Income and Wealth* 50(1): 87–106.
- Calderón, C., and L. Servén. 2014. The Effects of Infrastructure Development on Growth and Income Distribution. *Annals of Economics and Finance* 15(2): 521–534.
- Canning, D., and P. Pedroni. 1999. The Contribution of Infrastructure to Aggregate Output. Policy Research Working Paper 2246. Washington, DC: World Bank.
- Canning, D., and P. Pedroni. 2004. The Effect of Infrastructure on Long Run Economic Growth. Mimeo. Williamstown, MA: Williams College.
- Cronin, F. J., E. B. Parker, E. K. Collieran, and M. A. Gold. 1991. Telecommunications Infrastructure and Economic Growth: An Analysis of Causality. *Telecommunications Policy* 15(6): 529–535.
- Demetriades, P. O., and T. P. Mamuneas. 2000. Intertemporal Output and Employment Effects of Public Infrastructure Capital: Evidence from 12 OECD Economies. *The Economic Journal* 110(465): 687–712.
- Dercon, S. Risk, Poverty and Vulnerability in Africa. 2005. *Journal of African Economies* 14(4): 483–488.
- Du, Q., S.-J. Wei, and P. Xie. 2013. Roads and the Real Exchange Rate. NBER Working Paper No. 19291. Cambridge, MA: National Bureau of Economic Research.
- Easterly, W., and S. Rebelo. Fiscal Policy and Economic Growth. 1993. *Journal of Monetary Economics* 32(3): 417–458.

---

\* The Asian Development Bank refers to China by the name People's Republic of China.

- Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1998. An Analysis of Sample Attrition in Panel Data. *Journal of Human Resources* 33: 251–299.
- Gibson, J., and S. Rozelle. 2003. Poverty and Access to Roads in Papua New Guinea. *Economic Development and Cultural Change* 52(1): 159–185.
- Gramlich, E. M. 1994. Infrastructure Investment: A Review Essay. *Journal of Economic Literature* 32(3): 1176–1196.
- Hulten, C. R., E. Bennathan, and S. Srinivasan. 2006. Infrastructure, Externalities, and Economic Development: A Study of the Indian Manufacturing Industry. *The World Bank Economic Review* 20(2): 291–308.
- Moccerro, D. Improving the Business and Investment Climate in Indonesia. 2008. OECD Economics Department Working Papers, No. 638. Paris: OECD Publishing.
- Morrison, C. J., and A. E. Schwartz. 1996. Public Infrastructure, Private Input Demand, and Economic Performance in New England Manufacturing. *Journal of Business and Economic Statistics* 14(1): 91–101.
- National Bureau of Statistics of China. *China Statistical Yearbook*. Various years. China Statistics Press.
- Nerlove, M. 1967. Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections. *Economic Studies Quarterly* 18: 42–74.
- Piketty, T. 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.
- Röller, L.-H., and L. Waverman. 2001. Telecommunications Infrastructure and Economic Development: A Simultaneous Approach. *American Economic Review* 91(4): 909–923.
- Shorrocks, A., and G. Wan. 2005. Spatial Decomposition of Inequality. *Journal of Economic Geography* 5(1): 59–81.
- Straub, S. 2008. Infrastructure and Growth in Developing Countries. Policy Research Working Paper 4460. Washington, DC: World Bank.
- Wan, G. 2004. Accounting for Income Inequality in Rural China: A Regression-based Approach. *Journal of Comparative Economics* 32(2): 348–363.
- Wan, G. 2007. Understanding Regional Poverty and Inequality Trends in China: Methodological Issues and Empirical Findings. *Review of Income and Wealth* 53(1): 25–34.
- Wan, G., ed. 2008a. *Inequality and Growth in Modern China*. Oxford: Oxford University Press.
- Wan, G., ed. 2008b. *Understanding Inequality and Poverty in China: Methods and Applications*. Palgrave Macmillan.
- Wang, C., G. Wan, and D. Yang. 2014. Income Inequality in the People's Republic Of China: Trends, Determinants, and Proposed Remedies. *Journal of Economic Surveys* 28(4): 686–708.
- World Bank. 1994. *World Development Report*. Washington, DC: World Bank.