Wages and Access to International Markets: Evidence from Urban China

He, Xiaobo

School of Economics, University of Adelaide

23 February 2013

Wages and Access to International Markets: Evidence from Urban China *

Xiaobo He†

University of Adelaide

February 23, 2013

Abstract

Using China Household Income Project Survey (2002) data, this paper addresses the causal relationship between individual wages and access to international markets. The ordinary least squares estimates show statistically insignificant and quantitatively zero effects of accessibility to international markets proxied by the length of contemporary transport routes connecting the origin city and its nearest major seaport. However, using prefecture-level population density in 1820 as exogenous variation in current transport routes, the two-stage least squares regressions provide an opposite picture indicating that every 1 percent increase in distance from the origin city to international markets (i.e. the nearest seaport), *ceteris paribus*, has a negative impact on individual wages of 0.086 percent. This causal effect remains robust to various sensitivity tests which include current labor market structure, historical factor endowments and initial population development.

JEL Classification: O12, O15, F16

Keywords: Market Access, Wages, Instrumental Variable

*I am grateful to Nadya Baryshnikova, Tatyana Chesnokova, Christopher Findlay, Faqin Lin, Stephanie McWhinnie, Nicholas C.S. Sim, Rong Zhu and seminar participants at University of Adelaide for their helpful comments. The usual disclaimer applies.

†Corresponding information: School of Economics, University of Adelaide, 10 Pulteney Street, Adelaide SA 5005, Australia; e-mail: xiaobo.he@adelaide.edu.au.

1 Introduction

In development economics literature, cross-country studies emphasise the contribution of transportation infrastructure to economic growth (e.g. Esfahani and Ramirez, 2003; Calderón and Chong, 2004; and Calderón and Servén, 2005), in which the positive link appears undebatable. However, China is a special case which makes the relationship between transportation infrastructure and economic growth more complicated. In China where economic growth was rapid and infrastructure investment was substantial during 1978 to 2008,[3] the increasing inter-regional disparity of transportation infrastructure drove heterogeneous income growth and economic development.[4] While the empirical studies on China (e.g. Démurger (2001); Chen and Yao (2011)) using highly aggregated data have arrived at a consensus that transportation infrastructure has a strong impact on economic growth, recent work by Banerjee et al. (2012), using Chinese county-level data, shows that proximity to transportation network does not necessarily cause GDP per capita growth. These contrasting findings raise a new question about how transportation infrastructure influences income from a micro perspective. This is important, as highly aggregated data like provincial and national data may not explicitly capture individual attributes that affect individual wages.

In the growing literature which uses aggregated endowments to explain micro-level income differentials, researchers indicate that the differences in accessibility to transportation infrastructure, as a proxy for market access, may lead to income disparities (Banerjee et al., 2012; Hou and Emran, 2012). The market access reflects the trading activities between the origin location and its domestic and international trade partners. The seminal work in this area is Hering and Poncet (2010) which follows a New Economics Geography (NEG) framework. Using urban data from the China Household Income Project Survey (CHIPS, 1995), Hering and Poncet (2010) show that geographical heterogeneity of market access has

[3]For example, in 2008, the Chinese government announced a package that included committing 1) RMB 1.52 trillion for infrastructure projects that included road construction and to increase the total highway length by about 40 percent between 2010 and 2020, and 2) RMB 1.41 trillion in 2010 and 2011 for railroad construction and to raise the total operating length by about 33 percent between 2010 and 2015.

[4]See Démurger (2001) and Chen and Yao (2011) for the periods 1985–1998 and 1978–2006, respectively.

significant impacts on the disparities of individual wages. Building on Hering and Poncet (2010), Kamal et al. (2012) employ a similar theoretical framework but extend the empirical analysis by including urban samples from both CHIPS (1995) and CHIPS (2002). They find that the impact of access to markets on individual wages is heterogeneous among different types of workers. Using CHIPS (2002) data, Hou and Emran (2012) show that the accessibility to market, as measured by the distance from the origin place to its local business centres and major seaports, significantly affects rural household consumption levels.

The major challenge of using cross-sectional income data is that omitted unobservables cannot be easily purged, as cross-sectional data do not contain time-series information thus fixed effects estimations like in Démurger (2001) cannot be used to eliminate time-invariant unobservable factors. Methodologically, one may eliminate city-level fixed effects by introducing city dummies in this case. However, it is not appropriate either, because the variable of interest is itself a part of fixed effects which will be removed by dummy variables. Therefore, following the studies that use instrumental variables to ameliorate omitted biases caused by endogenous problems (see Hering and Poncet, 2010; Hou and Emran, 2012; and Kamal et al., 2012, among others), this paper proposes a new instrumental variable employing historical information to identify the exogenous variation in length of transport routes connecting the origin city and its nearest major seaport.

The instrumental variable relies on historical information, i.e. China's inter-prefecture population in 1820. It is worth noting that an instrument including historical population does not violate the exclusion restriction only if the population is predetermined prior to the construction of the first railroad. The instrument proposed in this paper satisfies this condition, as the first Chinese railroad was completed in 1881, so it could not affect the distribution of population in 1820. In other words, the instrument is not correlated with unobservable error terms that may affect the endogenous variable (transport routes) and the dependent variable (individual wages).

Using urban samples of CHIPS (2002), this paper finds that instruments that use information from the historical population are significantly correlated with the length of transport

routes. The strong effect is not confounded by omitting city characteristics (e.g. geographical and topographical conditions, contemporary population and political ranking) that might contaminate the impact of historical population on the present transportation infrastructure, since additional covariates, such as elevation, slope, latitude, longitude, current population density, administrative level and autonomous status are carefully taken into account in the regressions. Following a Mincer (1974) wage equation, this study shows that while the OLS estimates provide statistically insignificant and quantitatively zero effects of accessibility to international markets, proxied by length of transport routes, on individual hourly wages in urban China, the 2SLS estimates indicate a clear causal relationship that an increase in distance from the origin city to the international market (i.e. the nearest seaport), *ceteris paribus*, has a negative impact on individual wages in that origin city. Furthermore, the causal result remains strongly resilient in a battery of sensitivity tests.

The rest of the paper is organised as follows. Section 2 addresses the related literature and the Chinese economic history. Section 3 outlines the data and the methodology. Section 4 presents the main empirical findings and the results of robustness checks. Section 5 provides the concluding remarks.

2 Background

2.1 Related Literature

In this paper, the market access refers to importing and exporting activities between locations which are affected by heterogeneous transport routes linking the origin city with the trade partners. This study benefits from the conceptual framework initiated by the seminal studies on the NEG (Krugman, 1991; Krugman and Venables, 1995; Fujita et al., 1999). In Fujita et al. (1999), the market access of a location is defined as the sum of the market capacity (total trade potential) of surrounding locations that open up their borders. It is worth nothing that

the computation of market access requires a weighting scheme based on bilateral distance of the origin location and its trade partners. Given the definition of market access, the NEG theory models nominal wages in a location as a function of the market access, and predicts that wages are relatively high at business centres.[5] The prediction is important for this paper for two reasons. First, it indicates a possible direction of the impact of accessibility to markets on wages from theoretical perspectives. Second, it implies the necessity of eliminating effects from accessibility to domestic markets, if one aims to identify the specific contribution of accessibility to international markets to wage levels.

Recent empirical studies also shed light on the world's largest developing economy, China, emphasising its rapid market liberalisation. During the second phase of economic reform (1992–1997),[6] Poncet (2005) finds that the domestic market fragmentation became severe,[7] indicating that the domestic market integration in a developing country like China may not be consistently improving. Poncet (2003) confirms that China's 'Open Door' policy succeeded in opening up its borders to international markets but failed to reduce inter-provincial trade barriers. Built on these preliminary explorations, Hering and Poncet (2010) estimate the wage equation embedded in an NEG framework, addressing the effect of access to both domestic and international markets on individual wages.[8] Using CHIPS (1995) data, they claim that the heterogeneous geography of market access has a significant impact on individual differences in economic returns to labour. In addition, they find greater wage sensitivity to market access for highly-skilled workers relative to low-skilled workers; and for workers in non-public firms compared with workers in public firms. Using both CHIPS (1995) and CHIPS (2002) data, Kamal et al. (2012) underscore that market access affects wages paid to both skilled and unskilled workers, and argue that this influence was stronger

[5] A business centre is the hub where trade quantity is the highest in the region. For example, a provincial capital can be the business centre within the province.

[6] The economic reform of China consists of three main phases, i.e. 1978–1991, 1992–1997 and 1998 onwards.

[7] Local authorities govern most economic activities in one region across different economic sectors. Thus, the Chinese market structure is characterised as a 'cellular' structure, which is also known as market fragmentation in China.

[8] This is the first attempt to estimate the Mincer (1974) wage equation in an NEG framework using micro data from China.

for workers in state-owned firms relative to workers in other types of firms (e.g. private firms) during 1995–2002.

Departing from Hering and Poncet (2010) and Kamal et al. (2012) who investigate urban wage earners, Hou and Emran (2012) look at rural households, and quantify the direct impact of being relatively far away from both domestic and international markets. Their findings, relying on CHIPS (1995), show that the better the access to markets, measured as the length of transport routes that connect households and domestic markets (local business centres) and international markets (the nearest seaports), the greater the positive impact on rural consumption, which supplements Hering and Poncet (2010) and Kamal et al. (2012). This paper accepts the measure of market access from Hou and Emran (2012), but emphasises the effect of access to international markets rather than general access to markets (e.g. Hering and Poncet, 2010; Kamal et al., 2012) on urban wage earners.

This study also extends those explorations in economic history that emphasise the importance of transportation infrastructure in the early stages of market integration and economic development across countries, for instance, Keller and Shiue (2008), Michaels (2008), Donaldson (2010) and Banerjee et al. (2012). Importantly, these studies have raised the issue of endogeneity of the initial construction of transportation infrastructure. A typically endogeneity problem could be, for example, the construction of highways. The initial design of highway routes is likely to correlated with local attributes such as population and income level which may affect the economic outcomes that researchers are investigating. Looking at the US in the 1950s, Michaels (2008) uses the original plan of routes proposed in 1944 as an instrument for highway locations to deal with this endogeneity issue. Relying on Chinese county-level data, Banerjee et al. (2012) use the distance of a county from the transport networks that connect historical treaty ports and business centres to construct an exogenous proxy of accessibility to current transportation infrastructure.[9] Donaldson (2010) studies the

[9]The treaty ports are Chinese territories which were conquered by the westerners after the First Opium War (1840). These ports were forced to open up to foreign trade during the second half of the 19th century to the early 20th century by unequal treaties. The Chinese government considered these treaties as unequal "because they were not negotiated by nations treating each other as equals but were imposed on China after a war, and because they encroached upon China's sovereign rights ... which reduced her to semicolonial

effects of railroad construction in 19th century India, underlining the contribution of railroads to initial spatial agglomeration which reduced inter-regional trade costs. Keller and Shiue (2008) show that the effects of steam trains in 19th century Europe were substantially larger than customs liberalisation and currency agreements in terms of increasing market size. This paper's venture into Chinese history aiming to find an appropriate instrument is informed by studies noted above.

Finally, the current studies that look at infrastructure development in China usually use aggregated data, for example, provincial data. Using provincial panel data (1985–1998), Démurger (2001) investigates how heterogeneous provision of transportation infrastructure could drive provincial income disparities in China. Relying on Chinese provincial panel data over the period 1978–2006, Chen and Yao (2011) find that the increasing proportion of the government budget allocated to investments in transportation reduced household consumption. Although the existing literature has found aggregated effects of transportation infrastructure on income and consumption, it is still possible to improve the understanding of the role of transportation infrastructure by introducing microeconometric analysis. In general, increased provision of transportation infrastructure may raise regional income growth, but it is not clear-cut whether this may have heterogeneous effects on the wage distribution. Thus, the empirical section of the paper intends to quantify the causal impact of the provision of transportation infrastructure, as measured by length of current transport routes that connect the origin city and its nearest seaport, on income inequality in urban China. This paper contributes to the literature by offering micro-level evidence built on individual income (wages only) data from urban China in 2002.

status"(Hsü, 1970). For example, the treaties were imposed the British (e.g. Treaty of Nanjing, Treaty of the Bogue, Convention of Peking and Boxer Protocol), the French (e.g. Treaty of Whampoa, Treaty of Tientsin, Convention of Peking and Boxer Protocol), the American (e.g. Treaty of Wanghia, Treaty of Tientsin and Boxer Protocol), and the German (Treaty of Tientsin and Boxer Protocol), among others.

2.2 Historical Factor Endowments

The central concern of this paper is to find exogenous variations in transport routes connecting a location and its nearest seaport in 2002 in China. Thus, this section discusses historical population at prefecture level in 1820 which serves as the instrument variable. Moreover, the effects on individual wages imposed by other potential historical endowments are documented as well.

Cao (2000) provides a rich dataset including prefecture-level population density in the Qing Dynasty (1776, 1820, 1851, 1880 and 1910). This study gathers the information from the 1820 'census' to construct the instruments.[10] Three main reasons for choosing this dataset are presented. First, the earliest railroad in China had been completed in 1881, implying that data from 1880 and 1910 'censuses' are not usable, as they could be affected by the determinants of initial railroads and thus this would invalidate exclusion restrictions in a 2SLS approach. Second, this study does not employ data in 1851 not only because the timing is close to the First Opium War (1840) but also because it is influenced by the Taiping Rebellion (1850–1864) both of which caused a severe decline of population across Chinese prefectures. Last, this study prefers the 1820 'census' to the 1776 'census', as the former is relatively more accurate.

The intuitions of using prefecture-level historical population density as instruments for current transport routes are presented as follows. First, the example in the 19th century US reveals a positive link between population and expanding demand for railroads (Fogel, 1962), which in part explains why current routes which are built on initial railway networks are designed according to the pattern of population distribution. Recall that prefecture population in 1820 was not affected by the existence of the modern railway system, which does not violate exclusion restrictions.[11] Second, the history of Chinese railroads which is reviewed by Ma (1983) offers some evidence that rail routes are associated with surrounding

[10] In the Qing dynasty (1644–1911), there were no formal population censuses according to the modern-time statistical and demographical criterion. However, the Qing government had relatively complete records of registered prefecture-level population in some years, e.g. 1776, 1820, 1851, 1880 and 1910.

[11] The inclusion of other factors such as geographical and topographical conditions in the analysis will ensure that the estimation does not omit key determinants of current transport routes.

population. For example, the Beijing-Shenyang railroad (construction: 1881–1930) connects Beijing, Tianjin, Hebei province and Liaoning province which were all areas of high population density in the late 1800s. Figures 1 and 2, which illustrate the railway density by the end of the Qing Dynasty (1911)[12] and the average railway density during 1978–2008, look similar. This suggests that the initial level of provision of railroads completed a century ago is positively related to the level of provision of infrastructure in the contemporary Chinese railway system.

It is not easy to gather information for various historical endowments. Fortunately, the identities of treaty ports can provide rich information describing the differing levels of historical endowments cities contain. Generally, treaty ports were Chinese territories that were forced to open up to foreign trade during the late Qing Dynasty (1842–1911) by unequal treaties imposed by the westerners, for example, the British, the French, the American and the German. The majority of treaty ports were seaports and became famous metropolitan centres in the 20th century such as Hong Kong, Shanghai and Guangzhou.

In the literature, e.g. Jia (2012), researchers use an indicator, i.e. treaty port, to proxy for an aggregated historical endowment of a particular prefecture. This is because, first, accurate Chinese historical data at prefecture or county level are never easy to collect; second, it is reasonable to believe that a treaty-port membership implies multi-dimensional attributes which determine a region's modern-time economic development, and ultimately the average individual income level.

Using the difference-in-difference approach, Jia (2012) points out that population as a key measure of economic development grew around 30 percent faster in treaty ports which were opened up to international markets after the First Opium War (1840) than the rest of

[12]Since provincial boundaries do not remain constant from 1911 onwards, this study uses the current Chinese map for provincial boundaries and the length of railways in 1911 to compute the railway density in Figures 1 and 2. Based on information such as railway track lengths, the years when construction started, the years when tracks opened, and the locations of origin and terminal stations of both main and spur lines provided by Ma (1983), it is possible to obtain the total lengths of railroads that existed in 1911, which this study then assigns to each province according to contemporary provincial boundaries provided by *Google Maps* for the calculation of the historical railway density variable. In the calculation, this study eliminates the spur lines that were reported as dismantled, where the various reasons for the dismantling are discussed in Ma (1983).

regions that were not opened up during the second half of the 19th century. Furthermore, the effect from historical endowments, though often unobservable, plays a pivotal role in long-run economic growth. Findings from Jia (2012) show that treaty ports were among the first to integrate with international markets and took advantage of globalisation opportunities post to China's 'Open Door' policy implemented from 1978. On average, the growth rate of GDP per capita in treaty ports was 35 percent higher than in surrounding regions that were not forced to opened up to foreign trade after the First Opium War (1840), over the period 1988–2007 (Jia, 2012).

This paper emphasises the importance of the inclusion of historical endowments in the estimation of wage differentials between treaty ports and other cities. This is because they not only contribute information to contemporary size of population and GDP per capita growth by prefecture (city), but also because they have implications for the quality of institutions among Chinese cities, which in turn could drive economic divergence (Fang and Zhao, 2009). Fang and Zhao (2009) find that treaty ports are normally better in the sense of both economic growth and institutions. These institutions, i.e. social rules that protect property rights, can be interpreted as municipal authorities, police, and judiciaries established by westerners during the 1850s to the 1910s in treaty ports. Although one may argue about how tight the links between the 'western' institutions and today's Chinese socialism are, empirical findings from Fang and Zhao (2009) reveal that the large influence of the western institutions on China's transition from antiquity (prior to 1911) to modernity (1911 onwards) well explains contemporary institutional differentials among Chinese cities. This study therefore carefully controls for these historical endowments, proxied by an indicator for treaty ports, in the sensitivity tests to partial out unobservable confounding effects.

3 Data and Methodology

3.1 Data

This paper utilises individual information from the urban samples in the China Household Income Project Survey (CHIPS, 2002), covering 20,632 individuals and 6,835 households from 77 cities in 12 provinces.[13] Excluding those areas that do not have complete city-level information, the empirical analysis focuses on 63 cities and 7,979 urban wage earners aged from 18 to 60 whose personal characteristics of interest are not missing.[14] In order to visualise the geographical distribution of locations analysed in this paper, Figure 3 plots all sampled areas.

The dependent variable is the logarithm of individual hourly wage (lnW). In CHIPS (2002), individuals report their nominal annual income which consists of regular salaries, bonuses and subsidies (including housing, medical, child care and various regional subsidies). Following Hering and Poncet (2010), this study defines the hourly wage as total income per working hour. The actual working hours are obtained based on the number of claimed work days per month along with the number of working hours per day by individuals.[15] The causal variable of interest is the accessibility to international markets proxied by the logarithm of length of transport routes from a sampled location to its nearest major seaport (lnD). The three major seaports are Tianjin, Shanghai and Guangzhou located in the north-east, east and south-east, respectively. This study measures the transport routes as the shortest length of transport routes including railroads and highways which are provided by *Google Maps*.[16]

[13]These provinces (includes municipalities) are: Beijing, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Hubei, Hunan, Guangdong, Chongqing, Sichuan, Yunnan and Gansu. Note that this paper employs a broader definition of *city* which includes counties from the urban samples in CHIPS (2002) following Hering and Poncet (2010) and Kamal et al. (2012).

[14]This paper only looks at urban samples. For the sake of simplicity, the remaining sections refer to individual wages of urban wage earners as 'individual wages'.

[15]Note that this study has excluded the number of days that workers are unemployed.

[16]For remote counties and small cities that are located far from local business centres (provincial capitals), this study measures their accessibility to international markets as follows. First, the shortest highway route connecting the origin city and its provincial capital is recorded. Second, the shortest railway route connecting its provincial capital and one of the three major seaports is recorded. Last, the summation of the above two records is used as the distance to international markets for a remote location.

The summary statistics and definitions of full variables are provided in Table 1 and Table 8.

3.2 Methodology

The empirical model investigates the causal impact of access to international markets on the hourly wage for individual i from city j in province k as follows,

$$lnW_{ijk} = c + \phi lnD_{ijk} + \gamma' x_{ijk} + \beta' z_{jk} + u_{ijk} \tag{1}$$

The parameter c denotes the common intercept term across all individuals. lnW_{ijk} and lnD_{ijk} are the dependent and independent variables respectively. u_{ijk} is the disturbance. The causal effect of lnD_{ijk} on lnW_{ijk}, i.e. ϕ, explains to what extent the accessibility to international markets, as measured by length of current transport routes from the origin city to its nearest major seaport, could affect individual wages. Considering omitted variables that are correlated with individual wages, this study thus controls for two sets of covariates.

The first set of covariates (x_{ijk}) aims to pin down the unobserved individual heterogeneity. Following the Mincer (1974) wage equation, this study includes the information of individual age (including its square), working experience, gender and years of schooling in the empirical model. Additionally, as Li et al. (2007) among others point out, the identity of the Communist Party member could have an impact on income,[17] and this paper considers a binary indicator which is equal to unity if an individual is a member of the Chinese Communist Party. Given the fact that cultural and linguistic background could potentially affect income (see Gustafsson and Li, 2003; Gustafsson and Ding, 2009; Li and Ding, 2009), it is worth identifying whether an individual belongs to a minor ethnic group according to the official classification.[18] Following the literature (Zhao, 2002; Chen et al., 2005; Hering

[17]The positive selection effect of the Party membership exists at least for the older generation. Thus, the membership of the Communist Party is employed in the covariates as a safeguard.

[18]The latest population census (2010) reports that the proportion of *Han*, the only major ethnic group, is 91.51 percent (National Bureau of Statistics China, 2011). The remaining 55 groups are officially defined as minor ethnic groups in China.

and Poncet, 2010), this paper employs various indicators to partial out multi-dimensional fixed effects, for instance, sectoral categories, firms' ownerships, individual occupations and provincial unobservables.[19]

The second set of covariates is denoted by the vector z_{jk}, which consists of cities' attributes. The heterogeneous geographical and topographical conditions of locations are possibly associated with household income and consumption (Hering and Poncet, 2010; Hou and Emran, 2012), therefore, this paper controls for these confounding factors such as elevation, slope, latitude and longitude of each sampled city.[20] Moreover, the vector z_{jk} includes other characteristics of cities, for instance, current population density, administrative level and autonomous status, as they may be correlated with both individual wages and instrumental variables.[21] Omitting current population density is dangerous, because it is closely linked with historical population that is used to construct the instrumental variables used in the analysis. In other words, the failure to partial out this confounding effect invalidates the exclusion restriction that requires instruments to be uncorrelated with omitted variables. In addition, administrative level is important. It in part denotes the political ranking of a city within a province, which may determine the priority of development in the Chinese political hierarchy hence local income level. Meanwhile, it to some extent implies the historical development of a location, which may be influenced by the population density in the 1820s. Last, autonomous status[22] is included as it captures fundamental features of ethnically minor counties that are culturally and linguistically distinct from major Chinese *Han* regions.

Furthermore, the inclusion of cities' characteristics purges the potential impact of access to domestic markets on individual wages.[23] In practice, it is not easy to measure the in-

[19] All fixed effects are captured by dummy variables. See Table 9 and Table 10 for detailed classifications.

[20] These data are collected from `http://www.heavens-above.com`. See Table 8 for detailed definitions.

[21] This study also takes current labour market structure, historical endowments, and initial population into account in the robustness checks.

[22] Counties where minor ethnic groups are major residents may be officially designated as autonomous counties.

[23] For instance, this study controls for a city's population density in 2001 which is correlated with its consumption potential. Moreover, this study includes a city's latitude and longitude to pin down its relative location. The consumption potential and location information can help to reduce the potential spill-over effects from domestic markets.

fluences from domestic markets. Some studies like Hering and Poncet (2010) and Kamal et al. (2012) have defined access to domestic markets. They first compute market capacity for a province according to inter-province input-output tables, and then allocate it to each city within the province by the GDP share of each constituent city. Alternatively, Hou and Emran (2012) define access to domestic markets as the distance from each rural household to the nearest local business centre which most likely is a provincial capital. In this paper, the potential confounding effect of surrounding domestic markets has been pinned down by the inclusion of cities' attributes, for instance, geographical location, i.e. latitude and longitude. Even if the geographical location *per se* fails to capture a large share of spill-over effects from domestic markets nearby, the additional control variables, like local population density, ranking of political hierarchy and administrative status, can serve as a multi-dimensional indicator to mimic a city's accessibility to domestic markets.

3.3 Instrumental Variables

Even if the issue of omitted variables is carefully taken care of by ruling out city-level and province-level time-invariant heterogeneity, OLS estimates could still be inconsistent due to the endogeneity problems. This is because, first of all, individuals (or households) choose the place to live based on their own abilities which affect their earning potential. This is usually unobserved but may affect individual income. Second, China's transportation infrastructure developed unevenly across regions during 1978–2006 (Chen and Yao, 2011). Intuitively, the distribution of highways and railroads is rarely random, and it is likely to be positively associated with abundant natural endowments, higher income levels and future development potential. To mitigate the potential bias in OLS estimates which cannot purge out all city-level unobservables, this paper proposes a new identification strategy, which uses the historical population density along the route that connects each origin city and its nearest large seaport as instruments.

The historical population should not violate the exclusion restriction, otherwise it becomes a problematic instrument for at least two reasons. First, the literature has shown that the population level and distribution are affected by the presence of railroads in the past (Atack et al., 2009; Donaldson, 2010). Second, for the case of China, Kung and Li (2011) find that railways facilitated migration from the North China Plain (north) to Manchuria (north-east) in the early 1900s, which confirms that the historical distribution of population could have been influenced by the development of initial railway system. In this regard, using historical population as an instrument in the context of China is possible in principle but formidably challenging in practice, as this requires the information from the early 1800s, a period for which accurate prefecture-level population data are extremely rare. After a venture into history, the author chooses to use data from Cao (2000) which offers detailed population information for each prefecture in 1820. The novel dataset allows the construction of a set of instrumental variables consisting of four components — the standard deviation, maximum, minimum, and ratio of the maximum over the minimum of prefecture-level population density along the transport routes that link the origin cities with their nearest major seaports.[24]

There may remain a concern about the exogeneity of the instruments, and one may question why this study does not use historical county-level population information which seems to be more accurate in this case. The responses to the above two concerns are as follows. First, it is better to let the data speak. The first-stage results reported in section 4.3 show that the instruments are sharp and robust. Second, countless changes of administrative boundaries during the last two centuries make it almost impossible to use historical county-level population information as instruments. In contrast, most boundaries of prefectures remain constant in their administrative areas (Cao, 2000), facilitating the attempt to construct historical instruments in this study.

To sum up, having carefully considered other possible channels relating historical population to contemporary economic activities, this paper seeks to use the heterogeneity of

[24]Summary statistics are available in Table 1.

inter-prefecture population to generate the exogenous variation in accessibility to international markets which is proxied by length of transport routes connecting a city to its nearest major seaport.

4 Empirical Results

4.1 OLS Estimates

Table 2 summarises the compressed results from the OLS regressions. At first glance, the OLS estimates reveal that being far from international markets could significantly lower individual wages as column (1) shows. However, ignoring individual characteristics and city-level fixed effects, the OLS estimate tends to overemphasise the effect of *access to international markets* on individual wages. This effect attenuates to less than half when province fixed effects are included in column (2), suggesting that regional heterogeneity is crucial in explaining China's inter-regional wage differentials. Alternatively, one may interpret the importance of the inclusion of province fixed effects by looking at the R^2 in column (2) that solely explains around 6 percent of variations in urban wages.[25] In column (3), the estimation examines how individual heterogeneity affects wages. Surprisingly, the effect of accessibility to international markets on wages tapers off to almost zero and becomes statistically insignificant, which confirms that neglecting individual attributes may cause serious omitted variable bias. Moreover, the R^2 reported in column (3) shows that individual characteristics explain two more percentage points of total variations in wages than province dummies do.[26] Last, column (4) shows zero and insignificant impacts of *access to international markets* on individual wages when city characteristics and province dummies are considered only.

As the effect of access to international markets on urban wages becomes statistically

[25]It is derived from the exclusion of the 5 percent of variation which is explained by *access to international markets* in column (1).

[26]The explained fraction of wages variation by individual attributes is 13 percent. This is derived from $0.29 - 0.11 - 0.05 = 0.13$.

insignificant and quantitatively small in the baseline OLS estimation (column (4) of Table 2), further investigation intends to control for more covariates to eliminate biases due to omitted variables. Table 3 looks at the suitability of the OLS estimations with the inclusion of both individual characteristics and regional heterogeneity. As shown in columns (1) to (5), all coefficients of individual attributes show reasonable signs. For instance, the impact of age on wages is concave. Male urban wage earners tend to have an advantage compared to female urban wage earners in the Chinese labour market. Schooling, not surprisingly, plays an important role in determining individual wages. The results in Table 3 are consistent with the finding in Li et al. (2007) that being a Chinese Communist Party member is positively correlated with higher wages. But no significant evidence is found to support the argument that ethnic minorities suffer from lower wages. However, for the variable of interest, *access to international markets*, the OLS regressions show unexpected signs and insignificant coefficients, even after multi-dimensional fixed effects are soaked up by ownership, occupation and sector dummies in column (5). In other words, the inclusion of more covariates cannot easily ameliorate the omitted variables bias. Therefore, given the confidence that OLS estimates are troublesome, this study employs 2SLS regressions in the following discussions.

4.2 2SLS Estimates

Using the same covariates from Table 3, Table 4 reports the 2SLS estimates employing historical population information as instruments. Unlike the OLS estimates which are statistically insignificant and quantitatively close to zero, the 2SLS estimates paint a different picture, indicating a significantly negative causal impact of access to international markets on individual wages.

Generally, the causal effects become stronger with the increase in the number of multi-dimensional fixed effects considered in columns (1) to (5), implying the presence of heterogeneity among urban wage earners in China. This result is consistent with the findings in

Hering and Poncet (2010) who underline the within-province heterogeneous effects on individual wages. Interestingly, Table 4 reports similar coefficients of all regressors but the independent variable compared with that which have been found in Table 3, suggesting the endogeneity of *access to international markets*.

Using the estimate in column (5) as a benchmark, one can interpret that every 1 percent increase in distance from the origin city to the international market (i.e. the nearest seaport), *ceteris paribus*, has a negative impact on individual wages of 0.086 percent. The effect is smaller than that found by Hering and Poncet (2010) who report 0.136 as the elasticity of market access derived from a similar wage equation controlling for the same fixed effects.[27] This may be because this study uses transport routes as a proxy only for access to international markets,[28] while Hering and Poncet (2010) derive the indicator for both access to domestic and international markets from an NEG model.[29] Although the findings in this paper and in Hering and Poncet (2010) are not directly comparable, their results considered as a benchmark imply that at first pass the results of this study do not fall within an unreasonable range.

As a model-based assessment of instrument validity, this study employs the overidentifying restrictions test.[30] The Hansen J tests in Table 4 all well exceed 0.1 thus one cannot reject the null hypothesis, indicating that there is no evidence that the quantitatively significant estimates this study finds above are an unintended consequence of model misspecification and questionable instruments validity. Besides, this paper also reports the Kleibergen and

[27]See column (5) in Table 1 in Hering and Poncet (2010).

[28]The effects of domestic markets are probably ruled out by controlling for detailed city attributes including population density and administrative level, among others.

[29]Since this study does not estimate the same causal effect as Hering and Poncet (2010), it is hard to say that this study underestimates the impact of accessibility to international markets on individual wages in urban China. Moreover, even though this paper estimates a broader concept of market access, the estimates could still be smaller than those in Hering and Poncet for two reasons. First, this paper has carefully controlled for cities' characteristics in the wage equation but Hering and Poncet have not. Comparing R^2s in Table 4 with those in Table 1 in Hering and Poncet (2010), one can see that the variation in individual wages explained in this paper on average is twice as much as that which has been explained in Hering and Poncet. Second, as Hering and Poncet have used CHIPS (1995) data, the smaller estimates this paper reports could perhaps be the consequence of a convergence of economic returns to market access, which indicates that a catch-up effect occurred in inland China during 1995–2002.

[30]Rejecting the test implies that one or more instruments are invalid, or the model is misspecified, or both. But it does not distinguish which condition or if both conditions are violated.

Paap (KP) (2006) Wald statistic for evidence on instrument weakness.[31] As clearly shown in Table 4, all KP Wald statistics are sufficiently large so that it is not necessary to worry about the issue of weak instruments. To provide further evidence, Table 5 offers the first stage results where the overall F-statistics are also large enough to avoid standard weak instruments.

4.3 Robustness Checks

Using historical instruments in empirical studies is challenging, as they are likely to be correlated with omitted determinants of individual wages. For the sake of robustness, this section offers two sets of sensitivity tests, responding to those concerns that first, instrumental variables may be associated with macro-level unobservables and second, the causal effect found in section 4.2 may be driven by extreme values.

Table 6 reports the first set of robustness checks aiming to partial out other channels through which historical population density could potentially affect individual wages. Following Hering and Poncet (2010), the first robustness check includes city-level skill intensity which implies the accumulation of local human capital. The indicator, *skill intensity*, is defined as the proportion of the adult total labour force in the CHIPS (2002) that had completed at least nine years of schooling.[32] Accounting for skill intensity, this paper finds that the coefficient of *access to international markets* remains statistically significant but quantitatively smaller than in column (5) of Table 4. The shrink in the magnitude of the effect is possibly due to external returns to schooling[33] in Chinese cities as indicated in Liu (2007).[34]

[31]The KP statistic is evaluated against a critical value, adopted from Stock and Yogo (2005), that 10 percent is the maximum rejection rate the researcher is willing to tolerate if the true rejection rate is 5 percent. This follows from the suggestion of Stock et al. (2002).

[32]Nine years of schooling is equivalent to the completion of junior high school in China.

[33]External returns to schooling are the economic gains from the sharing of knowledge among workers. The key assumption is that a worker will gain a wage premium with the increase in the average level of schooling of his group (e.g. firm level, sectoral level and city level, among others).

[34]Using CHIPS (1988 and 1995), Liu (2007) finds a one-year increase in city average education could increase individual earnings by between 11 and 13 percent, which indicates the presence of external returns to schooling among Chinese cities.

The second robustness check looks at the inclusion of historical endowments proxied by a binary indicator which is equal to unity if a city was assigned to be a treaty port during the late 1800s.[35] Treaty ports generally developed better, because the westerners established municipal authorities, police, judiciaries, manufacturing, and infrastructure in these cities (Banerjee et al., 2012; Jia, 2012). Along with the 'better' institutions and earlier industrialisation, the accessibility to railroads may facilitate the economic prosperity in the majority of treaty ports, which could confound the effect of *access to international markets* on wages.[36] However, column (2) shows that being a treaty port does not significantly raise individual wages given that all other factors are carefully controlled for.[37] More importantly, the effect of *access to international markets* on wages is the same as column (5) of Table 4, suggesting that the 2SLS estimates are tight and robust.

The final two robustness checks focus on the impact of the size of city-level population in 1953 and the early average rate of population growth during 1953–1964.[38] It is possible that historical population is correlated with initial population and its growth, although the population density in 2001 has been considered in equation (1). Therefore, one may argue that the instrumental variables violate the exclusion restriction in a 2SLS estimation. But, as columns (3) and (4) of Table 6 have shown, the coefficients of *access to international markets* neither attenuate to zero nor become statistically insignificant, which indicates that the historical instruments are not invalid. The coefficients herein are similar to the baseline result in column (5) of Table 4, implying that the inclusion of early status and growth of population does not confound the explanatory power of the instruments.[39] Furthermore, only the initial size of population in 1953 raises wages in 2002 but not the early growth rate

[35]Relevant information is obtained from Jia (2012).

[36]Note that even though prefecture population in 1820 used in this paper is not determined by the status of treaty port after the First Opium War (1840), historical population still could be correlated with unobservables that raise the likelihood of being a treaty port.

[37]Jia (2012) analyses 57 treaty ports and finds a positive relationship between being a treaty port and contemporary income over the period 1987–2007. In this study, only eight sampled cities were treaty ports in the late 1800s.

[38]This is because 1953 and 1964 are the two population censuses after 1949 when the People's Republic of China was founded. County-level population information is provided by National Bureau of Statistics China (1988) in which the first and second population census (1953 and 1964 respectively) data are available.

[39]This is further confirmed by the first stage results listed in columns (3) and (4).

of population.

Although the omitted variables problem has been carefully examined in Table 6, one may still be worried about the possibility that extreme values can bias estimates. To this end, Table 7 provides a further test to evaluate the sensitivity of the baseline 2SLS estimates to the exclusion of extreme values in both dependent and independent variables. Columns (1) and (2) show that the coefficients of *access to international markets* remain unchanged compared with the result reported in column (5) of Table 4, excluding the top and bottom 10 percent urban wage earners in the baseline model. This finding is crucial as it indicates that the causal effect revealed in the 2SLS estimation is neither driven by the highest income nor the lowest income group. Furthermore, Table 7 examines how the exclusion of the nearest and farthest urban wage earners in terms of routes distance from one's location to the nearest major seaport affects the 2SLS estimates, which is motivated by Banerjee et al. (2012). As one can see in columns (3) and (4), the coefficients of interest are still significant, although their magnitude departs from the baseline result in column (5) of Table 4. The interpretation is two-fold. Firstly, the significant results again suggest the 2SLS estimates are robust and reliable. Secondly, on the one hand, not accounting for the farthest urban wage earners tends to weaken the effect of access to international markets on wages, on the other hand, excluding the nearest urban wage earners strengthens this effect. Overall, the differential magnitude of these impacts complements the basic average estimates in Table 4, which contributes to a comprehensive understanding of the heterogeneous effect of being located far from international markets. In sum, the results presented in Table 7 confirm the robustness of the baseline 2SLS estimates.

5 Conclusion

Focusing on urban China, this paper proposes a new identification strategy that uses heterogeneity of inter-prefecture population density in 1820 to identify the effect of accessibility

to international markets on individual wages in urban China. It finds that the length of contemporary transport routes connecting the origin city and its nearest major seaport is strongly influenced by differentials in historical population density, which suggests that the set of instruments this study constructs successfully captures independent variation in access to international markets. Using these instruments, this paper finds that every 1 percent increase in distance from the origin city to international markets (i.e. the nearest seaport), *ceteris paribus*, has a negative impact on individual wages of 0.086 percent. An alternative interpretation could be that every 1 percent reduction of length of transport routes from the origin city to its nearest seaport can raise individual wages by 0.086 percent. This causal effect is robust to the inclusion of a large number of additional control variables such as current labour market structure, historical factor endowments, and initial population development, to running separate regressions that exclude the 10 percent highest-income and lowest-income urban wage earners respectively, and to running separate regressions that get rid of the 10 percent nearest and farthest urban wage earners in terms of routes distance from one's location to the nearest major seaport respectively.

This paper makes two contributions. First, it explores the research area of using historical information as instrumental variables to mitigate endogeneity problems of contemporary variables. The effort of providing new historical instruments is related to the work in Banerjee et al. (2012) and Fang and Zhao (2009). Although this is not the first attempt to employ historical information as instrumental variables, it is one of the pioneering studies that use this approach to identify causal effects in the area of China's economic development.

Second, the main finding confirms that the accessibility to international markets has significant and heterogeneous effects on individual wages among Chinese cities. Relating to Démurger (2001) that confirms the contributions of the improvement of transportation infrastructure to economic growth using provincial data, this paper provides some micro-level evidence that the heterogeneity of transportation infrastructure, i.e. the varying length of transport routes connecting the origin city and its nearest major seaport, may result in income disparities in urban areas.

This paper also has important policy implications. While some existing studies have pointed out the 'positive' effect of improving transportation infrastructure which can raise regional economic growth, this paper reveals the 'negative' effect that it may widen the wage gap between urban wage earners. The presence of the negative effect indicates the drawback of a policy that focuses on raising investment in transportation infrastructure. Therefore, policy-makers should realise that reducing wage inequality in urban China while maintaining high rates of regional economic growth could be difficult to manage. A possible solution may require complementary policies that can efficiently deal with the increasing wage inequality among urban wage earners.

References

Atack, J., F. Bateman, M. Haines, and R. A. Margo (2009). Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850-60. Working Paper 14640, National Bureau of Economic Research.

Banerjee, A., E. Duflo, and N. Qian (2012). On the road: Access to transportation infrastructure and economic growth in China. Working Paper 17897, National Bureau of Economic Research, Inc.

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 (2005). The effects of infrastructure development on growth and income distribution. World Bank Policy Research Working Paper 3643, World Bank, Washington, D.C.

Cao, S. (2000). *Zhongguo Renkou Shi Qing shiqi (Population History of China, Qing Period)*, Volume 5. Shanghai: Fudan University Press.

Chen, B. and Y. Yao (2011). The cursed virtue: Government infrastructural investment

and household consumption in Chinese provinces. *Oxford Bulletin of Economics and Statistics 73*(6), 856–877.

Chen, Y., S. Démurger, and M. Fournier (2005). Earnings differentials and ownership structure in Chinese enterprises. *Economic Development and Cultural Change 53*(4), 933–958.

Démurger, S. (2001). Infrastructure development and economic growth: An explanation for regional disparities in China? *Journal of Comparative Economics 29*(1), 95–117.

Donaldson, D. (2010). Railroads of the Raj: Estimating the impact of transportation infrastructure. NBER Working Papers 16487, National Bureau of Economic Research, Inc.

Esfahani, H. S. and M. T. Ramirez (2003). Institutions, infrastructure, and economic growth. *Journal of Development Economics 70*(2), 443–477.

Fang, Y. and Y. Zhao (2009). Do institutions matter? Estimating the effect of institutions on economic performance in China. BOFIT Discussion Papers 9/2009, Bank of Finland, Institute for Economies in Transition.

Fogel, R. W. (1962). A quantitative approach to the study of railroads in American economic growth: A report of some preliminary findings. *Journal of Economic History 22*(2), 163–197.

Fujita, M., P. Krugman, and A. J. Venables (1999). *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: MIT Press.

Gustafsson, B. and S. Ding (2009). Villages where China's ethnic minorities live. *China Economic Review 20*(2), 193–207.

Gustafsson, B. and S. Li (2003). The ethnic minority-majority income gap in rural China during transition. *Economic Development and Cultural Change 51*(4), 805–822.

Hering, L. and S. Poncet (2010). Market access and individual wages: Evidence from China. *Review of Economics and Statistics 92*(1), 145–159.

Hou, Z. and M. S. Emran (2012). Access to markets and rural poverty: Evidence from household consumption in China. *Review of Economics and Statistics*. forthcoming.

Hsü, I. C.-Y. (1970). *The Rise of Modern China*. New York: Oxford University Press.

Jia, R. (2012). The legacies of forced freedom: China's treaty ports. *Review of Economics and Statistics*. conditional acceptance.

Kamal, F., M. E. Lovely, and P. Ouyang (2012). Does deeper integration enhance spatial advantages? Market access and wage growth in China. *International Review of Economics Finance 23*(C), 59–74.

Keller, W. and C. H. Shiue (2008). Institutions, technology, and trade. NBER Working Paper 13913, National Bureau of Economic Research, Inc.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*(1), 97–126.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy 99*(3), 483–499.

Krugman, P. R. and A. J. Venables (1995). Globalization and the inequality of nations. *Quarterly Journal of Economics 110*(4), 857–880.

Kung, J. K. S. and N. Li (2011). Commercialization as exogenous shocks: The effect of the soybean trade and migration in Manchurian villages, 1895-1934. *Explorations in Economic History 48*(4), 568–589.

Li, H., P. Liu, J. Zhang, and N. Ma (2007). Economic returns to communist party membership: Evidence from urban Chinese twins. *Economic Journal 117*(523), 1504–1520.

Li, S. and S. Ding (2009). An empirical analysis of income inequality between a minority and the majority in urban China: The case of Ningxia Hui autonomous region. Global COE

Hi-Stat Discussion Paper Series GD08-022, Institute of Economic Research, Hitotsubashi University.

Liu, Z. (2007). The external returns to education: Evidence from Chinese cities. *Journal of Urban Economics 61*(3), 542–564.

Ma, L. (1983). *Zhongguo tielu jianzhu biannian jianshi 1881-1981 (The chronological history of Chinese railroad construction 1881-1981)*. Beijing: Zhongguo Tiedao (China Railways) Press.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *Review of Economics and Statistics 90*(4), 683–701.

Mincer, J. A. (1974). *Schooling, Experience and Earnings*. New York: Columbia University Press.

National Bureau of Statistics China (1988). *China Population Statistics Yearbook (1988)*. Beijing: China Zhanwang Press.

National Bureau of Statistics China (2011). *The Sixth National Population Census Gazette (No.1)*.

Poncet, S. (2003). Measuring Chinese domestic and international integration. *China Economic Review 14*(1), 1–21.

Poncet, S. (2005). A fragmented China: Measure and determinants of China's domestic market disintegration. *Review of International Economics 13*(3), 409–430.

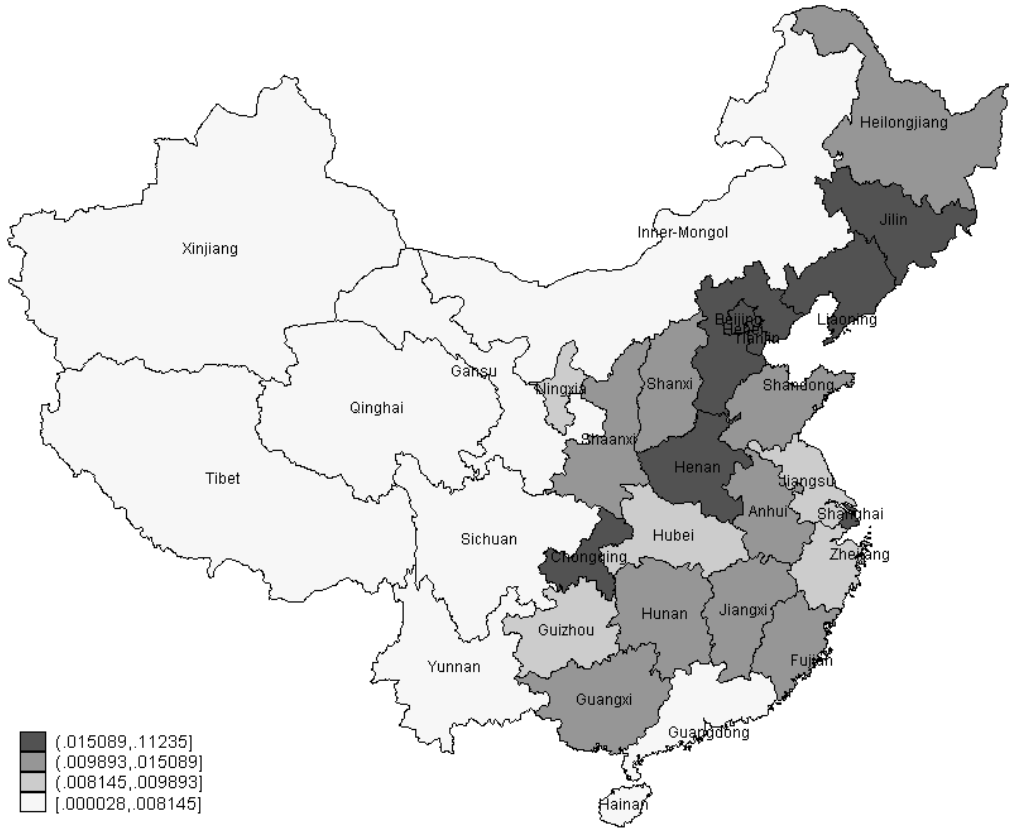
Stock, J. H., J. H. Wrigh, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*(4), 518–529.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. In D. W. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric*

Models: Essays in Honor of Thomas Rothenberg, Chapter 5, pp. 80–108. Cambridge: Cambridge University Press.

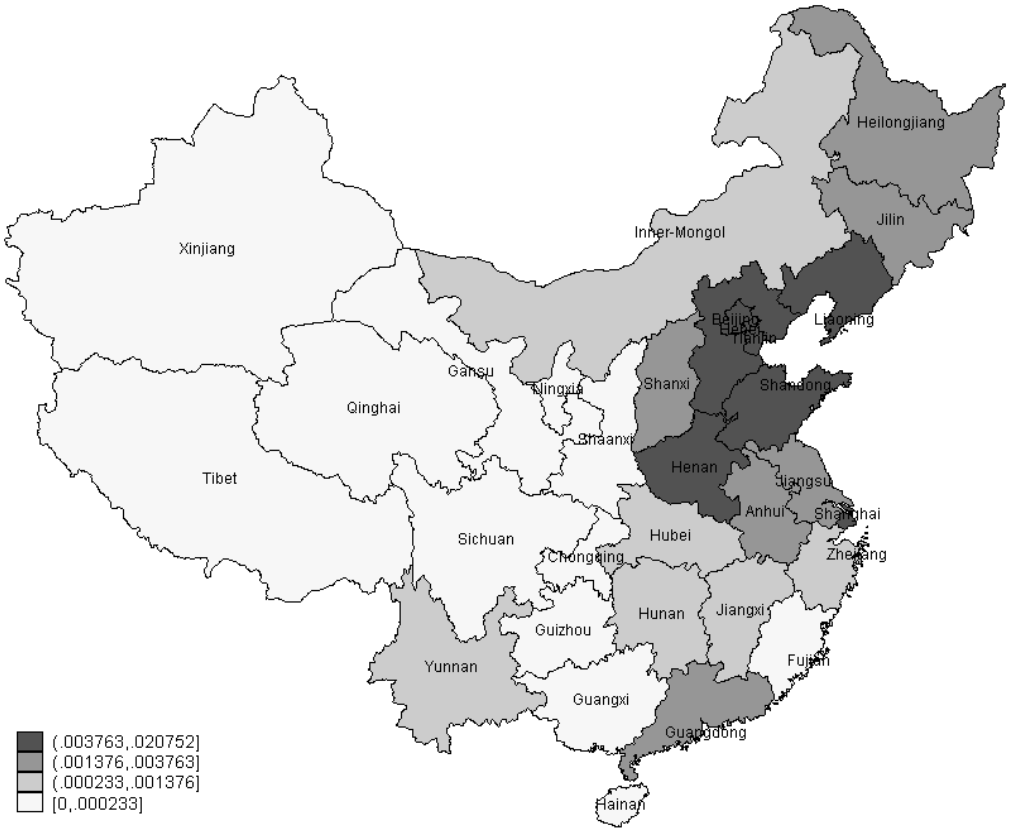
Zhao, Y. (2002). Earnings differentials between state and non-state enterprises in urban China. *Pacific Economic Review* 7(1), 181–197.

Figure 1: Average Railway Density (1978–2008)



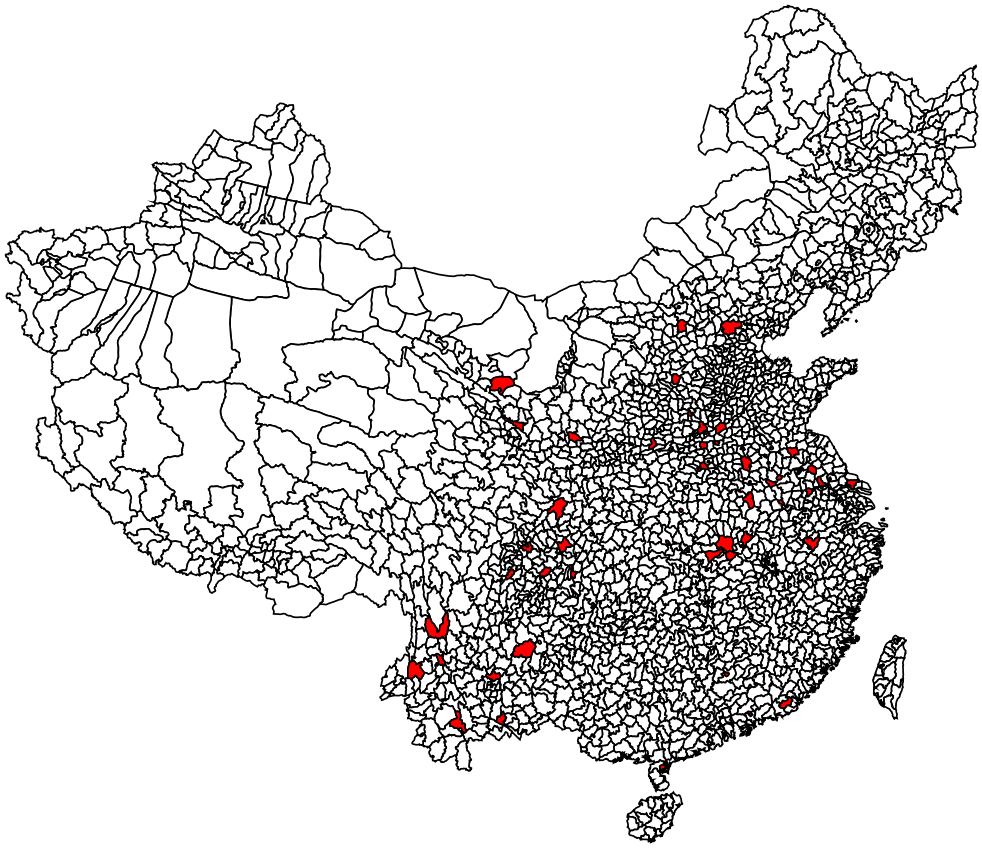
Source: National Bureau of Statistics of China (NBSC).

Figure 2: Railway Density (1911)



Source: Ma (1983).

Figure 3: Map of Sampled Areas



Source: China Household Income Project Survey (CHIPS, 2002) and National Bureau of Statistics of China (NBSC).

Table 1: Summary Statistics

Variables	Obs	Mean	SD	Min.	Max.
	Individual Characteristics				
log of hourly wage	7,979	1.49	0.71	-2.52	4.92
age	7,979	40.39	8.9	18	60
age^2	7,979	1,710.96	709.99	324	3,600
experience	7,979	20.07	9.54	0	41
male	7,979	0.56	0.50	0	1
schooling	7,979	11.46	3.01	0	23
communist party member	7,979	0.29	0.45	0	1
ethnic minority	7,979	0.04	0.20	0	1
	City Characteristics				
access to international markets	63	6.58	1.04	4.60	7.90
elevation	63	408.48	653.12	4	2395
slope	63	0.03	0.03	0.0003	0.12
latitude	63	32.23	5.17	21.2	41.12
longitude	63	110.92	11.90	29.58	121.14
population density (2001)	63	1.00	0.83	0.05	4.13
administrative level	63	0.24	0.43	0	1
autonomous regions	63	0.05	0.22	0	1
skill intensity	63	0.84	0.37	0	1
treaty ports	63	0.13	0.34	0	1
log of population (1953)	63	12.77	1.15	9.71	14.83
population growth (1953–1964)	63	0.04	0.04	-0.02	0.19
	Instrumental Variables				
population density along routes (sd)	63	104.45	76.78	13.93	293.47
population density along routes (max.)	63	409.95	263.44	102.3	874.1
population density along routes (min.)	63	102.51	104.06	10.4	531.7
population density along routes (max./min.)	63	6.40	4.53	1.15	20.34

Note: The definitions of variables are provided in Table 8.

Table 2: OLS Regressions 1

	(1)	(2)	(3)	(4)
	Dependent Variable: *log of hourly wage*			
access to international markets	-0.158***	-0.062**	-0.029	0.009
	(0.008)	(0.030)	(0.027)	(0.034)
Individual characteristics included?	No	No	Yes	No
City characteristics included?	No	No	No	Yes
Province dummies included?	No	Yes	Yes	Yes
Observations	7,979	7,979	7,979	7,979
R-squared	0.05	0.11	0.29	0.13

Note: The numbers in parentheses are robust standard errors. The definitions of variables are provided in Table 8.

Table 3: OLS Regressions 2

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable: *log of hourly wage*				
access to international markets	0.023	0.015	0.011	0.018	0.008
	(0.031)	(0.029)	(0.029)	(0.029)	(0.028)
	Individual Characteristics				
age	0.048***	0.053***	0.050***	0.052***	0.056***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
age^2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
experience	0.016***	0.013***	0.011***	0.013***	0.010***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
male	0.113***	0.119***	0.090***	0.103***	0.085***
	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)
schooling	0.072***	0.055***	0.045***	0.050***	0.034***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
communist party member	0.162***	0.114***	0.094***	0.125***	0.079***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
ethnic minority	-0.056*	-0.050	-0.035	-0.051	-0.044
	(0.033)	(0.034)	(0.032)	(0.032)	(0.032)
	City Characteristics				
elevation	0.000***	0.000**	0.000***	0.000***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
slope	0.202	0.709	-0.091	0.594	0.759
	(0.964)	(0.932)	(0.925)	(0.907)	(0.887)
latitude	-0.047***	-0.035***	-0.044***	-0.050***	-0.041***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
longitude	-0.001	-0.000	-0.002*	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
population density (2001)	0.026**	0.027**	0.018*	0.036***	0.026**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
administrative level	-0.146***	-0.113***	-0.162***	-0.177***	-0.149***
	(0.026)	(0.025)	(0.025)	(0.024)	(0.024)
autonomous regions	0.237***	0.264***	0.233***	0.236***	0.236***
	(0.049)	(0.048)	(0.047)	(0.047)	(0.046)
Province dummies included?	Yes	Yes	Yes	Yes	Yes
Ownership dummies included?	No	Yes	No	No	Yes
Occupation dummies included?	No	No	Yes	No	Yes
Sector dummies included?	No	No	No	Yes	Yes
Observations	7,979	7,979	7,979	7,979	7,979
R-squared	0.30	0.36	0.35	0.36	0.41

Note: The numbers in parentheses are robust standard errors. The definitions of variables are provided in Table 8.

Table 4: 2SLS Regressions (Second Stage)

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable: *log of hourly wage*				
access to international markets	-0.066*	-0.065*	-0.078**	-0.081**	-0.086***
	(0.035)	(0.033)	(0.034)	(0.034)	(0.033)
	Individual Characteristics				
age	0.049***	0.054***	0.050***	0.053***	0.057***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
age^2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
experience	0.016***	0.013***	0.011***	0.013***	0.010***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
male	0.113***	0.119***	0.090***	0.103***	0.085***
	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)
schooling	0.072***	0.055***	0.045***	0.050***	0.034***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
communist party member	0.164***	0.115***	0.095***	0.126***	0.079***
	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)
ethnic minority	-0.056*	-0.050	-0.034	-0.051	-0.044
	(0.033)	(0.034)	(0.032)	(0.032)	(0.032)
	City Characteristics				
elevation	0.000***	0.000**	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
slope	-0.649	-0.062	-0.934	-0.335	-0.133
	(0.977)	(0.943)	(0.935)	(0.917)	(0.896)
latitude	-0.038***	-0.027***	-0.034***	-0.040***	-0.031***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
longitude	-0.002	-0.000	-0.002**	-0.002*	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
population density (2001)	0.027**	0.028**	0.019*	0.037***	0.027**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
administrative level	-0.137***	-0.106***	-0.153***	-0.167***	-0.140***
	(0.026)	(0.025)	(0.025)	(0.025)	(0.024)
autonomous regions	0.231***	0.259***	0.226***	0.229***	0.229***
	(0.049)	(0.048)	(0.047)	(0.047)	(0.046)
Province dummies included?	Yes	Yes	Yes	Yes	Yes
Ownership dummy included?	No	Yes	No	No	Yes
Occupation dummy included?	No	No	Yes	No	Yes
Sector dummy included?	No	No	No	Yes	Yes
Weak IV	641.99	634.60	643.88	643.86	630.16
Hansen J	0.58	0.46	0.75	0.75	0.85
Observations	7,979	7,979	7,979	7,979	7,979
R-squared	0.30	0.35	0.35	0.37	0.40

Note: The numbers in parentheses are robust standard errors. *Weak IV* reports the Kleibergen-Paap Wald statistic. *Hansen J* reports the p-value for the overidentifying restriction test. The definitions of variables are provided in Table 8.

Table 5: 2SLS Regressions (First Stage)

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable: *access to international markets*				
population density along routes (sd)	-0.020***	-0.020***	-0.020***	-0.020***	-0.020***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (min.)	-0.007***	-0.007***	-0.007***	-0.007***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (max.)	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (max./min.)	0.014***	0.014***	0.014***	0.014***	0.014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Individual characteristics included?	Yes	Yes	Yes	Yes	Yes
City characteristics included?	Yes	Yes	Yes	Yes	Yes
Province dummies included?	Yes	Yes	Yes	Yes	Yes
Ownership dummies included?	No	Yes	No	No	Yes
Occupation dummies included?	No	No	Yes	No	Yes
Sector dummies included?	No	No	No	Yes	Yes
First-stage F-statistic	3.6×10^5	2.1×10^5	2.4×10^5	1.9×10^5	1.1×10^5
Observations	7,979	7,979	7,979	7,979	7,979

Note: The numbers in parentheses are robust standard errors. *F-statistic* reports the F-statistic for the overall significance of the first stage regressors. The definitions of variables are provided in Table 8.

Table 6: Robustness Checks

	(1)	(2)	(3)	(4)
	Dependent Variable: *log of hourly wage*			
access to international markets	-0.059*	-0.086***	-0.073**	-0.079**
	(0.032)	(0.033)	(0.032)	(0.033)
skill intensity	0.180***			
	(0.063)			
treaty ports		-0.016		
		(0.024)		
log of population (1953)			0.064***	
			(0.010)	
population growth (1953–1964)				-0.331
				(0.294)
	Dependent Variable: *access to international markets*			
population density along routes (sd)	-0.020***	-0.019***	-0.020***	-0.019***
	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (min.)	-0.007***	-0.007***	-0.007***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (max.)	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (max./min.)	0.014***	0.011***	0.017***	0.023***
	(0.000)	(0.002)	(0.002)	(0.002)
skill intensity	0.442***			
	(0.028)			
treaty ports		0.091***		
		(0.008)		
log of population (1953)			0.051***	
			(0.003)	
population growth (1953–1964)				1.018***
				(0.084)
First-stage F-statistic	60,221.69	1.1×10^5	1.5×10^5	94,462.59
Weak IV	1,271.99	662.41	635.24	668.18
Hansen J	0.52	0.84	0.68	0.93
Observations	7,979	7,979	7,979	7,979
R-squared	0.41	0.40	0.41	0.40

Note: All estimations include individual and city-level characteristics, province dummies, ownership dummies, occupation dummies and sector dummies. The numbers in parentheses are robust standard errors. *First-stage F-statistic* reports the F-statistic for the overall significance of the first stage regressors. *Weak IV* reports the Kleibergen-Paap Wald statistic. *Hansen J* reports the p-value for the overidentifying restriction test. The definitions of variables are provided in Table 8.

Table 7: Sample Sensitivity Test

	(1)	(2)	(3)	(4)
	Wage		Distance	
Omitted samples	Highest	Lowest	Farthest	Nearest
	Dependent Variable: *log of hourly wage*			
access to international markets	-0.081***	-0.074***	-0.067*	-0.167***
	(0.031)	(0.0289)	(0.039)	(0.041)
	Dependent Variable: *access to international markets*			
population density along routes (sd)	-0.020***	-0.021***	-0.019***	-0.015***
	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (min.)	-0.006***	-0.007***	-0.006***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (max.)	0.008***	0.008***	0.007***	0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
population density along routes (max./min.)	0.014***	0.014***	0.031***	0.001
	(0.000)	(0.002)	(0.002)	(0.001)
First-stage F-statistic	681,696.17	1.0×10^5	1.8×10^5	44138.81
Weak IV	550.46	574.06	824.62	623.66
Hansen J	0.20	0.72	0.75	0.08
Observations	7,182	7,181	7,292	6,951
R-squared	0.39	0.35	0.41	0.37

Note: Columns (1) and (2) exclude the top 10 percent and bottom 10 percent wage earners respectively. Columns (3) and (4) exclude the farthest 10 percent and nearest 10 percent inhabitants respectively. All estimations include individual and city-level characteristics, province dummies, ownership dummies, occupation dummies and sector dummies. The numbers in parentheses are robust standard errors. *First-stage* reports the F-statistic for the overall significance of the first stage regressors. *Weak IV* reports the Kleibergen-Paap Wald statistic. *Hansen J* reports the p-value for the overidentifying restriction test. The definitions of variables are provided in Table 8.

Table 8: List of Variables

Variable	Definition
	Individual Characteristics
log of hourly wage	the logarithm of individual hourly wage
age	age (18-60)
age^2	age squared
experience	the total length of working experience (years)
male	equal to 1 if male
communist party member	equal to 1 if a member of the Chinese Communist Party
ethnic minority	equal to 1 if belongs to a minor ethnic group
	City Characteristics
access to international markets	the logarithm of length of transport routes from a city to its nearest major seaport[1]
elevation	the elevation of a city where an individual is located (metre)
slope	the difference of elevation between a city and its nearest large seaport ÷ the linear distance between the two places
latitude	the latitude of a city where an individual is located
longitude	the longitude of a city where an individual is located
population density (2001)	the total population ÷ total area (1,000 people/km^2)
administrative level	equal to 1 if officially defined as a county, zero if city
autonomous regions	equal to 1 if an autonomous region due to ethnic minority
skill intensity	highly-skilled labour[2]÷total labour
treaty ports	equal to 1 if a historical treaty port post 1840
log of population (1953)	the logarithm of total population in 1953
population growth (1953–1964)	(total population in 1964÷total population in 1953)-1
	Instrumental Variables
population density[3] *along routes (sd)*	the standard deviation of population density along the route that connects a city to its nearest large seaport
population density along routes (max.)	the maximum value of population density along the route that connects a city to its nearest large seaport
population density along routes (min.)	the minimum value of population density along the route that connects a city to its nearest large seaport
population density along routes (max./min.)	the ratio of the maximum and minimum value of population density along the route that connects a city to its nearest large seaport

Note: [1] 'Major seaports' are defined as Tianjin, Shanghai and Guangzhou. [2] A 'highly-skilled' worker is one who completes nine or more years of education. [3] The population density is defined as total population ÷ total area (people/km^2).

Table 9: Category of Occupations

	Occupation
1	Owner (manager) of private firm
2	Self-employed
3	Professional
4	Director of government agent, institution
5	Department director of government agent
6	Clerical/office staff
7	Skilled worker
8	Unskilled worker
9	Sales clerk or service worker
10	Farmer
11	Other

Source: CHIPS (2002).

Table 10: Category of Sectors and Ownerships

	Sector	Ownership
1	Farming, forestry, husbandry and fishery	SOE at central/provincial level
2	Mining	Local SOE
3	Manufacturing	Urban collective
4	Electricity, gas and water supply facility	Private firm (including partnership)
5	Construction	Self-employed
6	Geological prospecting, irrigation administration	Sino-foreign joint venture
7	Transportation, storage, post office and communication	Foreign company
8	Wholesale, retail and food services	State shareholding company
9	Finance and insurance	Other shareholding company
10	Real estate	Rural private enterprise
11	Social services	Rural individual enterprise
12	Health, sports and social welfare	Others
13	Education, culture and arts, mass media and entertainment	
14	Scientific research and professional services	
15	Government agents, party organisations	
16	Others	

Source: CHIPS (2002).