

# Income per capita inequality in China: The Role of Economic Geography and Spatial Interactions

Laura Hering and Sandra Poncet<sup>\*†</sup>

## Abstract

This paper contributes to the analysis of growing income inequality in China. We apply a structural model of economic geography to data on per capita income over 190 Chinese cities between 1995 and 2002, and evaluate the extent to which market proximity and spatial dependence can explain the growing income inequality between Chinese cities. The econometric specification explicitly incorporates spatial dependence in the form of spatially-lagged per capita income. We show that the geography of market access and spatial dependence are significantly correlated with per capita income in China. Market access is particularly important in cities with smaller migration inflows, which is consistent with NEG theory, whereas spatially-lagged per capita income matters more in cities with greater immigration. We conclude that the positive impact of spatially-lagged income partly results from labor mobility between neighbors, so that spatial dependence reflects the influence of migration, knowledge transfers and increasing competition between cities.

*JEL* classification: E1, O1, O5, R1.

Keywords: Income inequality, Economic geography, Spatial dependence, China.

---

<sup>\*</sup>Corresponding author: Centre d'Economie de la Sorbonne (CES), Université Paris 1 and CEPPII. Address: Maison des Sciences Economiques, Bureau 405, 106 Boulevard de l'Hôpital 75013 PARIS. Email: sandra.poncet@univ-paris1.fr

<sup>†</sup>We would like to thank Miren Lafourcarde, Pierre-Philippe Combes, Mathieu Crozet, Agnès Bénassy-Quéré and two anonymous referees for their helpful comments and suggestions.

# 1 Introduction

Over the last two decades, China has benefited from unprecedented income growth, but at the price of large and increasing spatial income disparities (Meng et al., 2005). We see that regions with low per-capita income are predominantly found at the geographical periphery, while richer regions are located at the center. This core-periphery structure is consistent with the New Economic Geography (NEG) theory. This theory appeals to increasing returns to scale and transport costs to explain the agglomeration of economic activity (Krugman, 1991 and Krugman and Venables, 1995). One key determinant of the regional income level in NEG models is the spatial distribution of demand. Locations closer to consumer markets (i.e. with better “market access”) enjoy lower transport costs and have therefore a higher income (Fujita et al., 1999). This positive relationship between income and market access is modeled in the NEG “wage equation” and has first been confirmed empirically in a cross-country study by Redding and Venables (2003).

Recent empirical work on Chinese data highlights the role of economic geography in the explanation of domestic inequality and includes Lin (2005), Ma (2006), De Sousa and Poncet (2007) and Hering and Poncet (2009).

Most of this work on China relies on province-level data on market access and income to show that greater market proximity is associated with higher wages, but these findings have also been confirmed at the micro level (Hering and Poncet, 2009).

NEG theory further predicts that the correlation between wages and the demand for local production will be stronger in regions with less immigration. If demand for

products coming from a given region increases, new workers are needed to ensure a higher production. In case of low immigration, the region experiences a relative lack of workers and cannot satisfy the additional demand. Consequently, goods prices and in turn wages rise more in these regions.<sup>1</sup> This may produce spatial heterogeneity in the effect of market access on wages due to migration. This issue has only been tested indirectly in Hering and Poncet (2009) by allowing the impact of market access to vary between qualified and unqualified workers, with the latter being more likely to migrate.

A number of other criticisms can be addressed to the work discussed above. First, differences in endowments, policies and institutions across locations are not properly controlled since location fixed effects are not included.<sup>2</sup>

A second possible shortcoming is that each location is assumed to be an isolated entity. But individual geographical units are relatively well-integrated due to migration, inter-regional trade, and technology and knowledge spillovers, as well as institutions (Buettner, 1999), which produce spatial dependence between locations. This means that economic characteristics, such as income, may be correlated across localities.<sup>3</sup>

Spatial dependence is considered to be a powerful force in the convergence process

---

<sup>1</sup> The model considers labor as the only input and sets long-run profits equal to zero, so that any rise in prices translates into higher wages. However, similar results will pertain with different inputs or positive profits.

<sup>2</sup> Hering and Poncet (2009), who look at the impact of market access at the city level, account for fixed effects only at the more aggregated provincial level.

<sup>3</sup> Spatial dependence refers to the absence of independence between geographic observations, and is defined as the correlation of a variable across geographic units. Spatial dependence should not be confused with spatial heterogeneity, which occurs when parameters vary across countries or regions depending on their location.

(Rey and Montouri, 1999), so that its omission in estimations could result in serious misspecification (Abreu et al., 2005). This problem has been highlighted in the Chinese context by Ying (2003), who estimates output growth using provincial data over the 1978-1998 period. A number of analyses of foreign direct investment in China have revealed the importance of spatial dependence at the provincial (Cheung and Lin, 2004; Coughlin and Segev, 2000) and city levels (Madariaga and Poncet, 2007).

Even though spatial econometrics has received increasing attention over recent years, past research on the impact of market access on income (whether in China or elsewhere<sup>4</sup>) has mainly ignored these potential problems and consequently the resulting parameter estimates and statistical inference are open to criticism. Hanson (2005) and Mion (2004) were the first to address the issue of spatial dependence in a NEG framework, based on the Krugman-Helpman model. Fingleton (2006) uses the same theoretical model to test NEG theory against urban economics theory, showing that taking spatial dependence into account can render market potential insignificant when using data at a very fine geographical level.

The current paper contributes to the better understanding of the relationship between market access and spatial inequality in China, mitigating the shortcomings of the previous literature by relying on a panel data set covering 194 Chinese cities between 1995 and 2002. With data on a number of years for a considerable number of locations at a fine geographical level, we can introduce fixed effects by city into our regressions to control for scale economies and factor endowments.

---

<sup>4</sup>Estimations of the impact of market access on cross-country per capita income include Redding and Venables (2004), Head and Mayer (2006) and Breinlich (2006), amongst others.

One important contribution of this paper consists in asking whether the impact of market access depends on the intensity of immigration or other city-level characteristics.

By explicitly incorporating spatial dependence in the form of spatially-lagged per capita income, we ensure that the effect of market access is purged of any agglomeration effects, which allows us to draw a more precise picture of the spatial interaction between locations. Whereas in the spatial econometrics literature authors rarely search for the sources of spatial dependence, we here investigate the channels through which income per capita is affected by the income in neighboring cities, notably via migration.

Our results confirm that access to sources of demand is indeed important in shaping income dynamics in China. While spatial dependence between Chinese cities also significantly matters for the spatial distribution of income, including spatially-lagged income does not affect significantly our estimates of the effect of market access. We estimate the elasticity of city-level per capita income to market access to be 0.07. This figure is slightly lower than that of market access on wages from province-level data in De Sousa and Poncet (2007), and from individual data by Hering and Poncet (2009). Growing differences in trade costs and market size between Chinese cities will therefore lead to increasing income inequality.

To see whether labor supply in the form of internal immigration influences the impact of market access on wages we ask whether the relationship between market access and income holds across all cities equally, or whether it holds only for locations with low or high levels of immigration. Our results are very consistent with the NEG

model, which predicts that the relationship between market access and income will be weaker as migration rises. We find that doubling market access leads to a 11% rise in income in locations with low immigration, but only a 3% rise in locations with high immigration. Our results confirm those in De Sousa and Poncet (2007) and Hering and Poncet (2009): the further liberalization of internal migration may help to mitigate widening spatial income inequality fueled by the further opening of the country.

In order to explain the economic mechanisms behind the spatial dependence, we test whether spatially-lagged income has a greater impact when the city has higher intra-provincial immigration or is surrounded by well-developed infrastructure. If this is the case, then the mobility of individuals and the knowledge transfers associated with these two features may well be important channels through which proximity affects economic development.

This paper proceeds as follows. Section 2 outlines the theoretical framework from which the econometric specification is derived. Section 3 briefly discusses the role of spatial dependence and how it is taken into account in our estimations, and Section 4 presents the data and develops the empirical strategy. Section 5 discusses the results and Section 6 concludes.

## 2 Theoretical framework: geography and income level

The theoretical framework underlying our empirical analysis is that of a standard New Economic Geography model (Fujita et al., 1999) to which worker skill heterogeneity across regions is added (Head and Mayer, 2006).

The economy is composed of  $i = 1, \dots, R$  regions and two sectors: an agricultural sector ( $A$ ) and a manufacturing sector ( $M$ ), which is interpreted as a composite of manufacturing and service activities. The agricultural sector produces a homogeneous agricultural good, under constant returns and perfect competition. The manufacturing sector produces a large variety of differentiated goods, under increasing returns and imperfect competition.

### 2.1 Demand side

All consumers in region  $j$  share the same Cobb-Douglas preferences for the consumption of both types of goods (A and M):

$$U_j = M_j^\mu A_j^{1-\mu}, \quad 0 < \mu < 1, \quad (1)$$

where  $\mu$  denotes the expenditure share of manufactured goods.  $M_j$  is defined by a constant-elasticity-of-substitution (CES) sub-utility function of  $n_i$  varieties:

$$M_j = \sum_{i=1}^R \left( n_i q_{ij}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)}, \quad \sigma > 1, \quad (2)$$

where  $q_{ij}$  represents the demand by consumers in region  $j$  for a variety produced in region  $i$ , and  $\sigma$  is the elasticity of substitution. Given the expenditure of region  $j$  ( $E_j$ ) and the *c.i.f.* price of a variety produced in  $i$  and sold in  $j$  ( $p_{ij}$ ), the standard two-stage budgeting procedure yields the following CES demand  $q_{ij}$ :

$$q_{ij} = \mu p_{ij}^{-\sigma} G_j^{\sigma-1} E_j, \quad (3)$$

where  $G_j$  is the CES price index for manufactured goods, defined over the *c.i.f.* prices:

$$G_j = \left[ \sum_{i=1}^R n_i p_{ij}^{1-\sigma} \right]^{1/1-\sigma}. \quad (4)$$

## 2.2 Supply side

Transporting manufactured products from one region to another is costly. The iceberg transport technology assumes that  $p_{ij}$  is proportional to the mill price  $p_i$  and shipping costs  $T_{ij}$ , so that for every unit of good shipped abroad, only a fraction ( $\frac{1}{T_{ij}}$ ) arrives. Thus, the demand for a variety produced in  $i$  and sold in  $j$  shown in eq. (3) can be written as:

$$q_{ij} = \mu (p_i T_{ij})^{-\sigma} G_j^{\sigma-1} E_j. \quad (5)$$

To determine the total sales,  $q_i$ , of a representative firm in region  $i$  we sum sales across regions, given that total shipments to one region are  $T_{ij}$  times quantities consumed:

$$q_i = \mu \sum_{j=1}^R (p_i T_{ij})^{-\sigma} G_j^{\sigma-1} E_j T_{ij} = \mu p_i^{-\sigma} M A_i, \quad (6)$$

where

$$MA_i = \sum_{j=1}^R T_{ij}^{1-\sigma} G_j^{\sigma-1} E_j, \quad (7)$$

represents the market access of each exporting region  $i$  (Fujita et al., 1999). Each firm  $i$  earns profits  $\pi_i$ , assuming that the only input is labor:

$$\pi_i = p_i q_i - w_i \ell_i, \quad (8)$$

where  $w_i$  and  $\ell_i$  are the wage rate and labor demand for manufacturing workers respectively.<sup>5</sup> We follow Head and Mayer (2006) in taking worker skill heterogeneity into account.<sup>6</sup> We assume that labor demand,  $\ell$ , depends on both output,  $q$ , and workers' education,  $h$ , as follows:

$$\ell_i = (F + cq_i) \exp(-\rho h_i), \quad (9)$$

where  $F$  and  $c$  represent the fixed and marginal requirements in “effective” (education-adjusted) labor units. The parameter  $\rho$  measures the return to education and shows the percentage increase in productivity due to an increase in the average enrollment rate in higher education. Replacing (9) in (8) and maximizing profits yields the

---

<sup>5</sup>Perfect competition in the agricultural sector implies marginal-cost pricing, so that the price of the agricultural good  $p^A$  equals the wages of agricultural laborers  $w^A$ . We choose good A to be the numeraire, so that  $p^A = w^A = 1$ .

<sup>6</sup>The role of spatial differences in the skill composition of the work force as an explanation of the spatial wage distribution is analyzed in detail in Combes et al. (2008).

familiar mark-up pricing rule:

$$p_i = \frac{\sigma}{\sigma - 1} w_i c \exp(-\rho h_i), \quad (10)$$

for the varieties produced in region  $i$ . Given this pricing rule, profits are:

$$\pi_i = w_i \left[ cq_i \left( \frac{\exp(-\rho h_i)}{\sigma - 1} \right) - F \exp(-\rho h_i) \right]. \quad (11)$$

We assume that free entry and exit drive profits to zero. This implies that the equilibrium output of any firm is:

$$q^* = \frac{F(\sigma - 1)}{c}. \quad (12)$$

Using the demand function (6), the pricing rule (10) and equilibrium output (12), we can calculate the manufacturing wage when firms break even:

$$w_i = \frac{\sigma - 1}{\sigma c \exp(-\rho h_i)} \left[ \mu M A_i \frac{c}{F(\sigma - 1)} \right]^{1/\sigma} = \alpha [\mu M A_i]^{1/\sigma} \exp(\rho h_i). \quad (13)$$

Equation (13) relates location  $i$ 's income level to market access and education. This equation illustrates the two different ways in which a location can adjust to a shock, for example an increase in local demand,  $E$ : the price and the quantity adjustment (Head and Mayer, 2006). First, in the case of perfect factor mobility, the number of firms and workers may increase, which affects the price index  $G$ . In this case, the adjustment takes place inside  $MA = \sum_{j=1}^R T_{ij}^{1-\sigma} G_j^{\sigma-1} E_j$  since  $G$  compensates

for the change in  $E$  and total market access is unaffected (quantity adjustment). Thus in a context of full mobility of workers, wages should not depend on  $MA$ . Income inequality between cities does not derive from differences in their relative position to demand. Alternatively, in the case of no factor mobility, the number of firms and workers remains unchanged and the change in  $E$  induces  $MA = \sum_{j=1}^R T_{ij}^{1-\sigma} G_j^{\sigma-1} E_j$  to rise which in turn translates into a wage increase (price adjustment). Thus in the case of factor mobility restrictions, higher demand drives up prices, which is compensated by an increase in wages to ensure that the zero-profit condition holds.

In China, migration has long been severely restricted by a specific Chinese institution: the *hukou* system. The *hukou* is a system of household registration, forcing people to live and work in the place where they have an official registration. This system controls population and renders migration costly since local authorities can impose various hurdles to obtaining the necessary registration (Au and Henderson, 2006). We thus expect prices and wages to adjust following a change in demand so that city level income per capita is correlated to access to markets. Section 5.1 will confirm this prediction. In Section 5.2 we extend the empirical assessment and investigate the role of immigration on the impact of  $MA$ . As explained above, we anticipate that cities characterized by large migration inflows display a lower wage elasticity to market access. By contrast the price adjustment is expected to be stronger in cities with low immigration.

### 3 The role of spatial dependence

A considerable literature is devoted to the importance of spatial patterns at the sub-national level (see for example Abreu et al. (2005) for a survey of the literature on spatial factors in growth). Consequently, Chinese cities in our analysis are not treated as isolated geographical areas (Fingleton, 1999; Rey and Montouri, 1999), but it is rather assumed that the income of a Chinese city may be linked to its neighbors' incomes. The degree of these spatial interactions is assumed to follow Tobler's (1970) first law of geography: "everything is related to everything else, but near things are more related than distant things".

Spatial dependence can come from different sources. It can result from the omission of variables with a spatial dimension, such as climate, latitude or topology. Spatial dependence is also often generated by spillovers (such as technology externalities) due to the mobility of goods, workers or capital.

Econometrically, spatial dependence can take two forms.<sup>7</sup> The first is spatial autocorrelation. This describes how regional income per capita can be affected by a shock to income per capita in surrounding locations. That is to say, a shock in surrounding localities spills over through the error term. If spatial autocorrelation is erroneously ignored, standard statistical inferences will be invalid; however, the parameter estimates are unbiased.

In this paper, we adopt (following the diagnostic tests discussed in Section 4.3 and shown in Table A.1 of Appendix A) the spatial lag model. This is of particular interest

---

<sup>7</sup>See Anselin and Bera (1998) for an excellent introduction to spatial econometrics.

in testing theories of economic growth (Blonigen et al., 2007). In the spatial lag form, spatial dependence is captured by a term similar to a lagged dependent variable and is thus often referred to as spatial autoregression. Using standard notation, this type of regression model can be written as:  $y = \rho W y + \beta X + \epsilon$ , where  $y$  is a  $n$ -element vector of observations on the dependent variable,  $W$  is a  $n$  by  $n$  spatial-weighting matrix,  $X$  is a  $n$  by  $k$  matrix of  $k$  exogenous variables,  $\beta$  is a  $k$  element vector of coefficients,  $\rho$  is the spatial autoregressive coefficient that is assumed to lie between -1 and +1, and  $\epsilon$  is a  $n$ -element vector of error terms. The coefficient  $\rho$  measures how neighboring observations affect the dependent variable. Ignoring the spatial autoregressive term means leaving out a significant explanatory variable, so that the estimates of  $\beta$  are biased and all statistical inference is invalid.

## 4 Data and construction of variables

We here wish to evaluate the extent to which proximity to markets can explain growing income inequality within China. Section 4.1 describes the data set. Section 4.2 spells out how our main variable of interest, market access, is constructed, and Section 4.3 explains how spatial dependence is accounted for to ensure unbiased estimates.

### 4.1 Data

The data set comes mainly from two city-level sources: (1) the Urban Statistical Yearbook, various issues, published by China's State Statistical Bureau; and (2) Fifty Years of the Cities in New China: 1949-1998, also published by the State Statistical

Bureau.

To calculate the spatial lag variable and the city's market access for the eight years of our sample period (1995-2002), data on 199 cities is available. In the final regressions, five cities are dropped due to missing human capital data for all of the eight relevant years.<sup>8</sup> Our final data set covers 194 prefecture-level cities spread over the entire territory (except for the provinces of Qinghai, Xinjiang and Tibet) and consists of information on the urban part of these cities. Table A-3 in Appendix A lists the 199 cities by province.

Although the model provides predictions on nominal wages, data limitations forced us to rely on GDP per capita. The same proxy for wages has been used also by Redding and Venables (2004). The problem with the wage data is two-fold. First, wage data is measured based on a survey of staff and workers instead of the total employed population and is thus clearly over-estimated. Second, there are a lot of missing values in the series. GDP per capita, although imperfect, does not suffer from those problems. The natural logarithm of this variable is then used to calculate the spatial lag variable, as described in Section 4.3.

Our baseline specification contains a human capital variable, which we obtain by dividing the city's student enrollment in institutions of higher education by the city's total population.<sup>9</sup>

Our regressions further introduce three indicators to account for city-specific in-

---

<sup>8</sup> These cities are Hegang, Tongchuan, Guigang, Beihai and Yunfu.

<sup>9</sup> Institutions of higher education refer to establishments which have been set up according to government evaluation and approval procedures, enrolling high-school graduates and providing higher-education courses and training for senior professionals. These include full-time universities, colleges, and higher/further education institutes.

come determinants: the capital stock, the stock of foreign direct investments (FDI stock) and employment.

The city's capital stock is calculated following the standard approach using yearly investment flows  $I$  and a depreciation rate,  $\delta$ , of 5%. The formula is given by

$$K_t = K_{t-1}(1 - \delta) + I_t$$

where  $K_t = I_t$  for 1990.<sup>10</sup> The FDI stock is calculated in the same way. Both variables are expected to be positively correlated with income, since, for a given number of workers, greater capital stock implies higher productivity and therefore higher wages and income.

Employment data will be used to reflect the urban employment level, as this latter is known to be negatively correlated with regional wages and therefore incomes.

In order to investigate how the intensity of incoming migration affects the sensitivity of income to market access, we will differentiate between high and low migration cities based on city-level migration data, which comes from the 2000 Population Census.

Table A-2 in Appendix A provides summary statistics for our main variables of interest for the two extreme years of our sample 1995 and 2002, to demonstrate economic developments in Chinese cities.

Appendix B provides two figures that emphasize the large heterogeneity of income per capita, market access and spatial lag between Chinese cities as well as the

---

<sup>10</sup> The differences in capital endowment before 1990 are captured by the city fixed effects.

positive relationship between them. The third figure does not indicate a significant convergence of city level income per capita between 1995 and 2002.

## 4.2 Construction of market access

We compute the city-level market access as in Hering and Poncet (2009). This method follows the strategy pioneered by Redding and Venables (2004) that exploits the information from the estimation of bilateral trade via a gravity equation. The bilateral trade data used in our gravity equation consists of the intra-provincial, inter-provincial and international flows of Chinese provinces, as well as intranational and international flows of partners (see Appendix C for details of the data sources).

The estimated specification is derived as follows. Summing Equation (5) over all of the goods produced in location  $i$ , we obtain the total value of exports from  $i$  to  $j$ :

$$X_{ij} = \mu n_i (p_i T_{ij})^{1-\sigma} G_j^{\sigma-1} E_j = s_i \phi_{ij} m_j, \quad (14)$$

where  $n_i$  is the set of varieties produced in country  $i$ ,  $s_i$  measures the “supply capacity” of the exporting region,  $m_j = G_j^{\sigma-1} E_j$  the “market capacity” of region  $j$ , and  $\phi_{ij} = T_{ij}^{1-\sigma}$  the “freeness” of trade (Baldwin et al., 2003).<sup>11</sup>

Freeness of trade is assumed to depend on bilateral distances ( $dist_{ij}$ )<sup>12</sup> and a series of dummy variables which indicate whether provincial or foreign borders are crossed.

---

<sup>11</sup> The variable  $\phi_{ij} \in [0, 1]$  equals 1 when trade is free and 0 when trade is entirely eliminated due to high shipping costs.

<sup>12</sup> The internal distance of a Chinese province or a foreign country  $i$  is modeled as  $\frac{2}{3} \sqrt{area_{ii}/\pi}$ .

$$\phi_{ij} = dist_{ij}^{-\delta} \exp \left[ -\varphi B_{ij}^f - \varphi^* B_{ij}^{f*} + \psi Contig_{ij} - \vartheta B_{ij}^c + \xi B_{ij}^i + \zeta_{ij} \right], \quad (15)$$

where  $B_{ij}^f = 1$  if  $i$  and  $j$  are in two different countries with either  $i$  or  $j$  being China and 0 otherwise,  $B_{ij}^{f*} = 1$  if  $i$  and  $j$  are in two different countries with neither  $i$  nor  $j$  being China and 0 otherwise,  $Contig_{ij} = 1$  if the two different countries  $i$  and  $j$  are contiguous and 0 otherwise,  $B_{ij}^c = 1$  if  $i$  and  $j$  are two different Chinese provinces and 0 otherwise, and  $B_{ij}^i = 1$  if  $i = j$  denotes the same foreign country and 0 otherwise. The error term  $\zeta_{ij}$  captures the unmeasured determinants of trade freeness.

Substituting Equation (15) into (14), capturing unobserved exporter ( $\ln s_i$ ) and importer ( $\ln m_j$ ) country characteristics à la Redding and Venables (2004) with exporting and importing fixed effects ( $cty_i$  and  $ptn_j$ ), adding a time dimension and taking logs yields the following trade regression:

$$\begin{aligned} \ln X_{ijt} = & cty_{it} + ptn_{jt} - \delta_t \ln dist_{ijt} - \varphi_t B_{ijt}^f - \varphi_t^* B_{ijt}^{f*} \\ & + \psi_t Contig_{ijt} - \vartheta_t B_{ijt}^c + \xi_t B_{ijt}^i + \zeta_{ij} \end{aligned} \quad (16)$$

We estimate Equation (16) for the period 1995 to 2002, using our complete data set of trade, but allow the coefficients and fixed effects to vary across years. Recall from Equation (7) that market access for each year is defined as  $MA_i = \sum_{j=1}^R \phi_{ij} m_j = \sum_{j=1}^R \phi_{ij} \exp(ptn_j)$ , the trade cost-weighted sum of the market capacities of all partner

countries.<sup>13</sup>

To compute the market access of cities, we apply Head and Mayer's (2006) allocation rule, whereby the estimated market capacity  $m_j = G_j^{\sigma-1} E_j$  of province  $j$  is allocated to subunits (cities)  $c$  according to their share in province  $j$ 's economic activity. This allocation rule relies on two hypotheses. The first is homotheticity, so that the expenditure of city  $c$  is given by  $E_c = (y_c/y_j)E_j$ , where  $y_c/y_j$  is city  $c$ 's share of provincial GDP. The second is that  $G_j$ , the supply index, is approximately constant within provinces, i.e.  $G_c = G_j$  for all cities in  $j$ . These two assumptions together yield the market capacity of each city,  $m_c = (y_c/y_j)m_j$ .

The province-level market capacity ( $m_j$ ), is then allocated to cities in province  $j$  according to the GDP share of each constituent city  $c$ :

$$m_c = G_c^{\sigma-1} E_c = (y_c/y_j)m_j = (y_c/y_j)G_j^{\sigma-1} E_j = (y_c/y_j) \exp(p\tau n_j) \quad (17)$$

Note that while the lack of sub-provincial trade data forces us to choose an allocation rule for provincial competition-weighted expenditure,  $m$ , the other component of market access,  $\phi$ , uses genuine city-level information.

The market access of city  $c$  in province  $P$  then consists of four parts: local market access (intra-city demand); provincial market access (rest of the province); national market access (demand from other Chinese provinces); and world market access.

---

<sup>13</sup> For conciseness, estimates of this trade equation are not shown but are available on request. Our results are in line with those of De Sousa and Poncet (2007) and interested readers are referred to this paper for more details.

$$\begin{aligned}
\widehat{MA}_{ct} &= \widehat{\phi}_{cct} G_{ct}^{\sigma-1} E_{ct} + \sum_{k \in P} \widehat{\phi}_{ckt} \frac{y_{kt}}{\sum y_{kt}} G_{Pt}^{\sigma-1} E_{Pt} + \sum_{j \in C} \widehat{\phi}_{cjt} G_{jt}^{\sigma-1} E_{jt} \\
&\quad + \sum_{j \in F} \widehat{\phi}_{cjt} G_{jt}^{\sigma-1} E_{jt} \\
&= dist_{cc}^{-\widehat{\delta}_t} \times (y_{ct}/y_{Pt}) \exp(ptn_{Pt}) + \sum_{k \in P} dist_{ck}^{-\widehat{\delta}_t} \times \frac{y_{kt}}{\sum y_{kt}} \exp(ptn_{Pt}) \\
&\quad + \sum_{j \in C} dist_{cj}^{-\widehat{\delta}_t} \times \exp(\vartheta_t) \times \exp(ptn_{jt}) \quad (18) \\
&\quad + \sum_{j \in F} dist_{cj}^{-\widehat{\delta}_t} \times \exp(\widehat{\varphi}_t + \widehat{\psi}_t Contig_{cj}) \times \exp(ptn_{jt}),
\end{aligned}$$

where  $P$ ,  $C$  and  $F$  stand for the city's province, the rest of China and foreign countries, respectively. The parameters  $\widehat{\delta}_t$ ,  $\widehat{\vartheta}_t$ ,  $\widehat{\varphi}_t$  and  $\widehat{\psi}_t$ , as well as  $ptn_{jt}$ , are estimated in the trade equation, while  $dist_{cj}$  are the great circle distances between  $c$  and  $j$ .

### 4.3 Calculation of spatially-lagged income

In this paper, the spatial lag of income per capita is introduced to ensure that any positive significant impact of market access is not affected by the spatial correlation of observations. The robust Lagrange Multiplier tests suggest that spatial dependence in our data is probable. The tests reported in Table A-1 in Appendix A reject the null hypothesis of error autocorrelation while they do not reject the presence of a spatial autoregression at the 10% confidence level. We therefore consider a spatial lag model.

The construction of the model relies on the weight matrix  $W$ , which contains

information about the relative dependence between the cities in our sample. The literature suggests a number of alternative weighting methods. The most widely-used are based on contiguity and distance between localities, but differ in the particular functional form retained. As recommended by Anselin and Bera (1998) and Keller (2002), the elements of the matrix have to be exogenous<sup>14</sup>, otherwise the empirical model becomes highly non-linear. We choose a spatial weighting matrix  $W$  that depends exclusively on the geographical distance  $d_{cj}$  between cities  $c$  and  $j$ , since the exogeneity of distance is absolutely unambiguous. We use the inverse squared distance in order to reflect a gravity relation. The distance-based weights,  $w_{cj}$ , are thus defined as

$$w_{cj} = 0, \text{ if } i = j$$

$$w_{cj} = 1/d_{cj}^2, \text{ if } d_{cj} \leq 800$$

$$w_{cj} = 0, \text{ if } d_{cj} > 800$$

The distance of 800 km is the cut-off level above which interactions are assumed to be negligible. This is important since there must be a limit to the range of spatial dependence allowed by the spatial weights matrix (Abreu et al., 2005).<sup>15</sup> We will check that our results are robust to changes in the cut-off level.

The matrix  $W$  is then row-standardized (with  $w_{cj}^*$  being an element of the standardized weight matrix) as  $w_{cj}^* = w_{cj} / \sum_j w_{cj}$ , so that each row sums to one and each weight may be interpreted as the city's share in the total spatial effect.

Using the standardized weight matrix  $W$ , our spatially-lagged income variable,

---

<sup>14</sup> This condition is a prerequisite for the introduction of spatial econometrics.

<sup>15</sup> This is due to the asymptotic property required to obtain consistent estimates of the model parameters.

*spatial lag*, is then given by  $Wy_{ct} = \sum_{j \neq c} (y_{jt} w_{cj}^*)$ .<sup>16</sup>

## 5 Empirical estimation results

### 5.1 Benchmark estimates

Having calculated market access at the city level,  $MA_c$ , and the spatial lag of our dependent variable, we can now estimate our human-capital augmented version of the wage equation. Taking the natural logarithms of Equation (13), introducing a time dimension and controlling for time-invariant city effects,  $\eta_c$ , and common time effects,  $\lambda_t$ , yields the following estimation equation:

$$\ln y_{ct} = a + b \ln MA_{ct} + \rho \ln human\ capital_{ct} + \eta_c + \lambda_t + \epsilon_{ct} \quad (19)$$

Our benchmark estimates are obtained using panel regression techniques (city-level and year fixed effects). We report bootstrap standard errors to control for the potential econometric problem that arises from the two-step calculation of our market access variable. This variable is calculated from parameters that are themselves estimated with standard errors in an initial regression. As a consequence we verified that our results were robust to the correction of the biased standard errors by applying the “bootstrap” procedure to each of our regressions.<sup>17</sup>

Column 1 of Table 1 shows the estimates of Equation (19), which reveal a positive and significant impact of market access and human capital on per capita income. As

---

<sup>16</sup> While income per capita varies over time, the spatial weight matrix remains unchanged.

<sup>17</sup> Only Column 8 of of Table 1 does not display bootstrap standard errors.

discussed by Head and Mayer (2006), the intercept  $a$  depends on the input requirement coefficients  $F$  and  $c$ . These are likely to vary across cities and time due to differences in capital intensity. As such, from Column 2 onwards, we control for city-level capital stock and employment, whose estimated parameters are of the expected sign.

In Column 3, we introduce the spatially-lagged dependent variable, to account for any spatial dependence in China. As explained above, the spatial lag of income per capita  $y$  in city  $c$  corresponds to the sum of spatially-weighted values of  $y$  for the surrounding locations.

The results suggest that spatial dependence between Chinese cities is important but that it does not alter our estimates of the determinants of city income. The estimated coefficients on the other explanatory variables are little changed from those in column 1. Accounting for spatial dependence leads to an increase of one percentage point in the  $R^2$  statistic.

Before looking in detail at the impact of market access, we consider two robustness checks for our spatial lag variable. As described in Section 3, the variable in Column 3 relies on a cut-off of 800 km. The results with other cut-offs for the spatial lag are very similar to those in Column 3, as shown in Columns 4 and 5 which are based on cut-offs of 600 km and 1,200 km, respectively.

Our estimates of the elasticity of city-level per capita income to market access are robust to the control for spatial dependence in per capita income. This coefficient of 0.07 is slightly lower than those obtained in province-level data by De Sousa and

Poncet (2007), and in individual data by Hering and Poncet (2009).<sup>18</sup>

Growing differences in trade costs or market size between Chinese cities can therefore produce rising income inequality. Our benchmark estimates (Column 3) imply that the doubling of market access that occurred between 1995 and 2002 would be associated with a 7% rise in per capita income. The coefficient of 0.2 on the spatial lag further suggests that changes in this variable between 1995 and 2002 correspond to a rise of 20% in per capita income.

To ensure that our results are reliable, Column 6 introduces population density, as larger and/or denser cities should benefit more from knowledge spillovers between firms and workers, leading to greater worker productivity and incomes. The first five columns did not sufficiently control for this aspect, so significant market access could reflect a size effect caused by spillovers between firms. The results show that the impact of  $MA$  is slightly reduced when population density is controlled for; this does not however change the flavor of our results.

We have not as yet addressed the potential simultaneity problem. City fixed effects control for time-invariant omitted variables, but reverse causality remains an issue. Market access, as an explanatory variable, is a weighted sum of all potential expenditures, including local expenditures. These expenditures depend on income, raising the concern of reverse causality. Since a positive income shock will raise  $E$  and thus  $MA$ , we use a two-fold approach to test the robustness of our estimates: we first estimate our equation in first differences (Column 7); second, we instrument market access (Column 8). In Hering and Poncet (2009) market access is instrumented by a

---

<sup>18</sup>Head and Mayer (2006) find a similar value (0.1) on European data.

variable called centrality, which measures the distance of each city in the sample to the center of every inhabited  $1^\circ$  by  $1^\circ$  cell in the world population grid. Here, we appeal to panel data, so centrality is not a valid instrument for market access, as it does not vary over time. In Column 8, we thus resort to two instruments which are inspired by the GMM strategy, even though they significantly reduce our sample size: the first and second differences in market access. Hansen's J-test of overidentifying restrictions does not significantly reject the validity of our instruments. We also report the Cragg-Donald F-statistic, suggested by Stock and Yogo (2002) as a global test for weak instruments (i.e. it tests the null hypothesis that a given group of instruments is weak against the alternative that they are strong). Our instrument set is accepted as strong, since the Cragg-Donald F-statistic exceeds the critical value of 10% maximum bias of the IV estimator relative to OLS at the 5% confidence level.

The next step is to carry out the Durbin-Wu-Hausman test, which tests for the endogeneity of market access in an IV regression. Since this test does not reject the null hypothesis of exogeneity of market access (at the 10% confidence level), we report the OLS estimates since they are more efficient than IV estimates (Pagan, 1984). All of the test statistics are displayed at the bottom of Table 1.

## **5.2 The heterogeneous influence of market access**

One novel contribution of this paper is to test whether the relationship between market access and income depends on city characteristics. It is likely that the contribution of market access to income inequality in China is rooted in not only the heterogeneity of market access across cities but also the heterogenous impact of market access on

income, depending on city characteristics, and notably immigration intensity.

According to our theoretical model, in the case of quasi-infinite labor supply for the manufacturing sector, wages will respond only little to changes in demand from international and local markets. This relates to the two different mechanisms by which the local economy can adjust to a change in the demand for its goods: either quantitative adjustment with new workers filling positions to answer the additional demand, or, in the case of insufficient labor mobility, adjustment takes the form of a change in prices, so that income rises with market access.

To see whether the impact of market access depends on immigration intensity, we consider two sub-samples: high- and low-immigration cities. The immigration rate used to make this distinction is taken from the 2000 population census, and is calculated as the ratio of incoming population with household registration in another city or county of the same province or another province over the city's total population. For cities with above-median immigration the "High immigration" dummy equals 1; those below the median are classified as low immigration.

Column 1 interacts market access with this "High immigration" dummy, while columns 2 and 3 run separate regressions for high- and low-immigration cities. Our results are consistent with the predictions: the market-access coefficient is large and significant at the 1% level in cities with low immigration; in high-immigration cities, market access has a much smaller effect, as shown by the negative and significant coefficient on the interaction term in Column 1 and the coefficients on market access in Column 2 and 3. According to these estimates, a doubling of market access is associated with a 11% rise in income in low-immigration locations, compared to a

3% rise in high-immigration locations. Our results are in line with those in Hering and Poncet (2009), who noted, using individual wage data, a greater effect of market access on the wages of skilled workers. They argue that high-skilled workers are likely to benefit more from market access as they are less at risk from migrants, who are in the majority low skilled.

No such heterogeneity results when economic development (instead of migration) is used to differentiate cities.<sup>19</sup> Given the insignificant interaction term of market access and the level of income per capita in Column 4, the impact of market access does not seem to be significantly different between these two groups.

### 5.3 What lies behind spatial dependence?

So far, we have only established that spatial dependence plays a role in determining the spatial distribution of income within China. The next step is to understand what lies behind this effect.

We have already taken the demand side into account via market access, so the significant spatial lag effect must reflect something other than pure demand. Two potential candidates are technology and knowledge spillovers. To check this hypothesis, we see whether the impact of spatial lags is stronger in environments which are more favorable to the exchange of ideas and factors.

We use two proxies for the ease of communication between neighboring cities: the rate of intra-provincial migration and the quality of the surrounding infrastructure. As in the previous section, we create two dummy variables, “High internal migration”

---

<sup>19</sup> We choose income per capita in 1990 as the criterion to split the sample.

and “Good surrounding infrastructure”, that we interact with the spatial lag.

As before, we use migration data from the 2000 population census, which distinguishes between migrants coming from the same province and those coming from other provinces. High internal immigration cities are those where the percentage of population coming from a different location within the same province is above the median.

To see whether a city is surrounded by good infrastructure we calculate the spatial lag of the variable “density of streets”.<sup>20</sup> “Good surrounding infrastructure” cities have values of this spatial lag of infrastructure above the median.

Both variables, intra-provincial migration and the spatial lag of infrastructure, reflect the mobility of factors and goods between closely located cities. We imagine that spatial dependence may have a stronger impact in regions with greater labor mobility. First, with greater mobility, wages in one location will move in line with wages in surrounding cities, to avoid losing workers to neighbors. Second, migration creates spillovers. Migrants from more productive cities bring knowledge with them which may well improve productivity in the immigration location. Last, a city that is known to accept migrants and offer job opportunities might attract qualified migrants and consequently increase productivity and income.

A well-developed infrastructure facilitates communication and commuting between cities and therefore knowledge transfers and spillovers. Fewer impediments to the exchange of factors and goods will intensify the competition between cities and

---

<sup>20</sup> As for the spatially-lagged income variable, we weigh the density of streets (with respect to population) of each neighboring city by the bilateral distance.

may bid up the price of labor in order to attract the required work force.

We thus expect the interaction terms, introduced in Columns 5 and 6, to be positive, reinforcing the spatial lag variable. This is indeed the case. The interactions of the spatial lag with both “High internal migration” and “Good surrounding infrastructure” are positive and significant.

The last column includes both interactions of the spatial lag as well as the interaction of market access with “High Immigration”. All three of the interactions are significant. The coefficient of the original spatially-lagged income is now much reduced. The impact of neighboring cities is therefore at least partly due to migration and its associated spillovers.

## 6 Conclusion

This paper has examined the role of economic geography and spatial dependence in explaining the spatial structure of per capita income in China. Our econometric specification relates city-level per capita income to a transport-cost weighted sum of demand in surrounding locations after controlling for spatial dependence and endowments. The data come from a sample of 194 Chinese cities between 1995 and 2002. We find that per capita income has increased due to both better market access and the reinforcement of spatial interdependence between Chinese cities.

This effect of market access on income inequality depends on local immigration intensity. The elasticity of income to market access is much higher in cities located in provinces which are characterized by lower immigration. While further international

trade integration of China is expected to fuel an upward pressure on wages, this can be mitigated by lower barriers to internal migration. In the light of this complementarity, further liberalization of internal migration may help to maintain the Chinese price competitiveness. These results confirm those in previous work (De Sousa and Poncet, 2007; Hering and Poncet, 2009).

We also find heterogeneity in the impact of spatially-lagged income, which has a stronger influence in cities with a higher percentage of intra-provincial migrants or which are surrounded by good infrastructure. This suggests that neighboring cities have a greater impact when the mobility of factors and goods is facilitated. In this case, knowledge spreads more easily and competition for workers is fiercer, increasing the city's income.

## 7 References

- Abreu M., H. De Groot, and R. Florax, 2005, "Space and Growth: A Survey of Empirical Evidence and Methods", *Région et Développement* 21, 13-44.
- Anselin L., and A. Bera, 1998, "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics", in *Handbook of Applied Economic Statistics*, Ullah A, Giles DEA (eds). Berlin: Springer-Verlag.
- Au C.-C., and V. Henderson, 2006, "How Migration Restrictions Limit Agglomeration and Productivity in China," *Journal of Development Economics*, 80:2, 350-388.
- Baldwin R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud, 2003, *Economic Geography and Public Policy*, Princeton: Princeton University Press.

- Blonigen B. A., R. B. Davies, G. R. Waddell, and H. Naughton, 2007, "FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment", *European Economic Review* 51:5, 1303-1325.
- Breinlich H., 2006, "The Spatial Income Structure in the European Union - What Role for Economic Geography?", *Journal of Economic Geography* 6:5, 593-617.
- Buettner T., 1999, "The effects of unemployment, aggregate wages, and spatial contiguity on local wages: An investigation with German district level data", *Papers in Regional Science* 78, 47-67.
- Cheung K. and P. Lin, 2004, "Spillover Effects of FDI on Innovation in China: Evidence from the Provincial Data", *China Economic Review* 15:1, 25-44.
- Combes P.-P., G. Duranton and L. Gobillon, 2008, "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, 63:2, 723-742.
- Coughlin C. C., and E. Segev, 2000, "Foreign Direct Investment in China: A Spatial Econometric Study", *The World Economy* 23:1, 1-23.
- De Sousa J., and S. Poncet, 2007, "How are wages set in Beijing?", *CEPII, working paper*, 2007-13.
- Hanson G. H., 2005, "Market Potential, Increasing Returns, and Geographic Concentration", *Journal of International Economics* 67:1, 1-24.
- Head K., and T. Mayer, 2006, "Regional Wage and Employment Responses to Market Potential in the EU," *Regional Science and Urban Economics*, 36(5), 573-594.
- Hering L. and S. Poncet, 2009, "Market access and individual wages: evidence from China", *Review of Economics and Statistics*, forthcoming.
- Fingleton B., 1999, "Estimates of Time to Economic Convergence: An Analysis of Regions of the European Union", *International Regional Science Review* 22(1): 5-34.
- Fingleton B., 2006, "The new economic geography versus urban economics: an evaluation using local wage rates in Great Britain", *Oxford Economic Papers* 56(3): 501-530.
- Fujita M., P. Krugman, and A. J. Venables, 1999, *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, Cambridge MA.
- Keller W., 2002, "Geographic Localization of International Technology Diffusion", *American Economic Review* 92: 120-142.
- Krugman P., 1991, "Increasing Returns and Economic Geography", *Journal of Political Economy*, 99(3), 483-499.
- Krugman P. and A. J. Venables, 1995, "Globalization and the Inequality of Nations", *Quarterly Journal of Economics*, 110(4), 857-880.
- Lin S., 2005, "Geographic Location, Trade and Income Inequality in China", in R. Kabur and A. J. Venables Spatial (eds.) *Inequality and Development*, Oxford University Press, London.

- Madariaga, N. and S. Poncet, 2007, "FDI in China: spillovers and impact on growth", *The World Economy* 30, 837-862.
- Ma A. C., 2006, "Geographical Location of Foreign Direct Investment and Wage Inequality in China", *The World Economy* 29:8, 1031-1055.
- Meng X., R. Gregory, and Y. Wang, 2005, "Poverty, Inequality and Growth in Urban China, 1986-2000", *Journal of Comparative Economics* 33:4, 710-729.
- Mion G., 2004, "Spatial externalities and empirical analysis: the case of Italy", *Journal of Urban Economics* 56, 97-118.
- Pagan A., 1984, "Model Evaluation by Variable Addition", in D.F. Hendry and K.F. Wallis (eds) *Econometrics and Quantitative Economics*, Blackwell: Oxford, chapter 5, 103-134.
- Redding S. J., and A. J. Venables, 2004, "Economic Geography and International Inequality", *Journal of International Economics* 62:1, 53-82.
- Rey S., and B. Montouri, 1999, "U.S. Regional Income Convergence: A Spatial Econometric Perspective", *Regional Studies* 33, 145-156.
- Stock J. H. and M. Yogo, 2002, "Testing for weak instruments in linear IV regression", NBER technical working paper 284.
- Tobler W. R., 1970, "A Computer movie simulating urban growth in the Detroit Region", *Economic Geography* 46, 234-240.
- Wei S.-J., 1996, "Intra-National Versus International Trade: How Stubborn Are Nations in Global Integration", *NBER Working Paper* 5531.
- Ying L. G., 2003, "Understanding China's recent growth experience: A spatial econometric perspective", *The Annals of Regional Science*, 37:4, 613-628.

Table 1: Benchmark estimations. Dependent variable: per capita income

	(1) All	(2) All	(3) All	(4) 600 km	(5) 1200 km	(6) All	(7) 1st Diff	(8) IV
Market Access	0.09*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.06*** (0.02)	0.05*** (0.02)	0.06*** (0.02)
Human capital	0.09*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.06*** (0.02)	0.08*** (0.02)	0.08*** (0.02)
Capital stock		0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.08*** (0.03)	0.03 (0.02)	0.01 (0.03)
FDI stock		0.06*** (0.02)	0.06*** (0.02)	0.06*** (0.01)	0.06*** (0.02)	0.05*** (0.01)	0.04** (0.02)	0.09*** (0.02)
Employment		-0.05*** (0.01)	-0.05*** (0.01)	-0.05*** (0.01)	-0.05*** (0.01)	-0.03*** (0.01)	-0.04*** (0.01)	-0.05*** (0.01)
Spatial lag			0.21** (0.10)	0.21*** (0.09)	0.23** (0.10)	0.24** (0.10)	0.15** (0.07)	0.28*** (0.08)
Density						0.14*** (0.02)		
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	no	Yes
Observations	1437	1437	1437	1437	1437	1437	1241	1081
$R^2$	0.75	0.76	0.76	0.76	0.76	0.78	0.16	0.68
Number of cities	194	194	194	194	194	194		188
Hansen J statistic (Prob>F)								1.07 (0.3)
Durbin-Wu-Hausman stat (Prob>F)								2.2 (0.14)
Cragg-Donald F stat								529.94
Critical value (10%)								19.93

Bootstrap heteroskedastic-consistent standard errors in parentheses, with \*\*\*, \*\* and \* denoting significance at the 1, 5 and 10% levels respectively. Standard errors are corrected for clustering at the province level. The reported R-squared is the Within R-squared, which indicates how much of the variation of wages within the group of sectors is explained by our regressors. The critical value for the Cragg-Donald test is based on a 10% 2SLS bias at the 5% significance level (see Stock and Yogo, 2002).

Table 2: Heterogeneity depending on city characteristics

	(1) All	(2) High immig.	(3) Low immig.	(4) All	(5) All	(6) All	(7) All
Market Access	0.11*** (0.03)	0.03* (0.02)	0.11*** (0.03)	0.08*** (0.03)	0.07*** (0.02)	0.07*** (0.02)	0.11*** (0.03)
MA*“High Immigration”	-0.06** (0.03)						-0.07** (0.03)
MA*“High income per capita”				-0.01 (0.03)			
Human capital	0.08*** (0.02)	0.10*** (0.02)	0.07*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)
Capital stock	0.05* (0.03)	0.07* (0.04)	0.02 (0.03)	0.05 (0.03)	0.05* (0.03)	0.05* (0.03)	0.06** (0.03)
FDI stock	0.06*** (0.02)	0.11*** (0.02)	0.01 (0.02)	0.05*** (0.02)	0.05*** (0.02)	0.05*** (0.02)	0.05*** (0.02)
Employment	-0.05*** (0.01)	-0.04* (0.02)	-0.07*** (0.02)	-0.05*** (0.01)	-0.06*** (0.01)	-0.05*** (0.01)	-0.06*** (0.01)
Spatial lag	0.21** (0.10)	0.18* (0.11)	0.22 (0.14)	0.21** (0.10)	0.17* (0.09)	0.16 (0.11)	0.10 (0.11)
Sp. lag*“High internal immig.”					0.09** (0.04)		0.09*** (0.04)
Sp. lag*“Good surr. infrastr.”						0.08** (0.04)	0.09** (0.04)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1437	716	721	1437	1437	1437	1437
$R^2$	0.76	0.82	0.72	0.76	0.76	0.76	0.77
Number of cities	194	96	98	194	194	194	194

Bootstrap heteroskedastic-consistent standard errors in parentheses, with \*\*\*, \*\* and \* denoting significance at the 1, 5 and 10% levels respectively. Standard errors are corrected for clustering at the province level. The reported R-squared is the Within R-squared, which indicates how much of the variation of wages within the group of sectors is explained by our regressors.

## Appendix A

Table A-1: Spatial dependance diagnostics

Test	Statistic	p-value
<i>Spatial error:</i>		
Moran's I	-0.307	1.241
Lagrange multiplier	0.564	0.453
Robust Lagrange multiplier	0.002	0.960
<i>Spatial lag:</i>		
Lagrange multiplier	7.967	0.005
Robust Lagrange multiplier	7.406	0.007

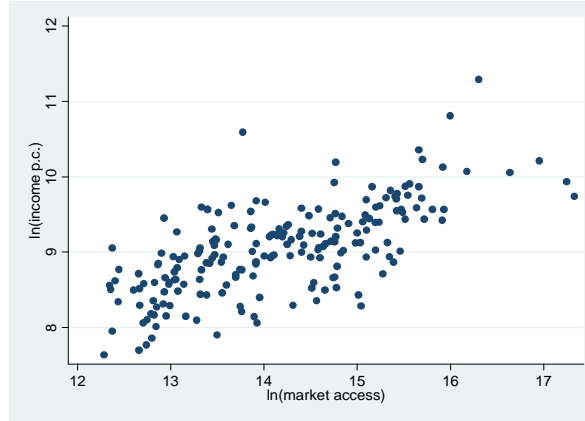
Table A-2: Summary statistics

Variable	Year	Mean	Std. Deviation
Market access (10,000 units)	1995	2,631	4,287
	2002	4,847	9,974
Per capita GDP (Yuan)	1995	10,414	7,841
	2002	18,559	15,084
Spatial lag	1995	9.07	0.33
	2002	9.64	0.31
Human capital (%)	1995	1.06	0.97
	2002	2.53	2.12
Capital stock (10,000 Yuan)	1995	1,166,501	2,325,257
	2002	5,287,426	1.16e+07
FDI stock (10,000 USD)	1995	36,305	81,539
	2002	134,674	351,92
Employment (10,000 persons)	1995	669.1	764.2
	2002	281	478.6

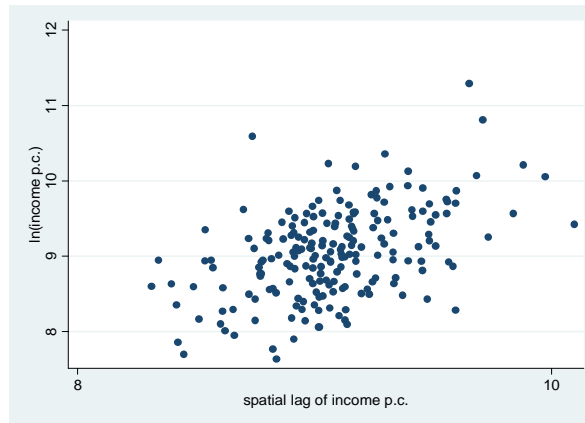
**Appendix A Table A-3: List of Cities**

Province	City
Beijing	Beijing
Tianjin	Tianjin
Hebei	Shijiazhuang, Baoding, Cangzhou, Chengde, Handan, Hengshui, Langfang, Qinhuangdao, Tangshan Xingtai, Zhangjiakou,
Shanxi	Taiyuan
Inner Mongolia	Baotou, Chifeng, Hohhot
Liaoning	Anshan, Chaoyang, Dalian, Dandong, Fushun, Fuxin, Huludao, Jinzhou, Liaoyang, Panjin Shenyang, Yingkou
Jilin	Baicheng, Baishan, Changchun, Jilin, Liaoyuan, Siping, Tonghua
Heilongjiang	Daqing, Gigihai, Harbin, Hegang, Heihe, Jiamusi, Jixi, Mudanjiang, Qitaihe Shuangyashan, Yichun
Shanghai	Shanghai
Jiangsu	Changzhou, Huaiyin, Liayungang, Nanjing, Nantong, Suzhou, Taizhou, Wuxi, Xuzhou Yancheng, Yangzhou, Zhenjiang
Zhejiang	Hangzhou, Huzhou, Jiaxing, Jinhua, Ningbo, Quzhou, Shaoxing, Wenzhou, Zhoushan
Anhui	Anqing, Bengbu, Chuzhou, Hefei, Huaibei, Huainan, Huangshan, Maanshan, Tongling, Wuhu
Fujian	Fuzhou, Longyan, Nanping, Putian, Quanzhou, Sanming, Xiamen, Zhangzhou,
Jiangxi	Jingdezhen, Jiujiang, Nanchang, Pingxiang, Xinyu, Yingtan
Shandong	Dezhou, Dongying, Jinan, Jining, Laiwu, Linyi, Qingdao, Rizhao, Taian, Weifang, Weihai Yantai, Zaozhuang Zibo
Henan	Anyang, Hebi, Jiaozuo, Kaifeng, Luohe, Luoyang, Nanyang, Pingdingshan, Puyang, Sanmenxia Shangqiu, Xinxiang, Xuchang Zhengzhou
Hubei	Ezhou, Huanggang, Huangshi, Jingmen, Jingzhou, Shiyan, Wuhan, Xiangfan, Xiaogan
Hunan	Changde, Changsha, Chenzhou, Hengyang, Huaihua, Shaoyang, Xiangtan, Yiyang, Yongzhou Yueyang, Zhangjiajie, Zhuzhou
Guangdong	Chaozhou, Dongguan, Foshan, Guangzhou, Heyuan, Huizhou, Jiangmen, Jieyang, Maoming Meizhou, Qingyuan, Shantou, Shanwei, Shaoguang, Shenzhen, Yangjiang, Yunfu, Zhanjiang Zhaoqing, Zhongshan, Zhuhai
Guangxi	Beihai, Guigang, Guilin, Liuzhou, Nanning, Qinzhou, Wuzhou, Yulin
Hainan	Haikou, Sanya,
Chongqing	Chongqing
Sichuan	Chengdu, Deyang, Guangyuan, Leshan, Luzhou, Mianyang, Nanchong, Neijiang, Suining Yibin, Zigong
Guizhou	Guiyang, Luipanzhui, Zunyi
Yunnan	Kunming, Qujing,
Shaanxi	Baoji, Hanzhong, Tongchuan, Weinan, Xian, Xianyang, Yanan
Gansu	Lanzhou
Ningxia	Yinchuan

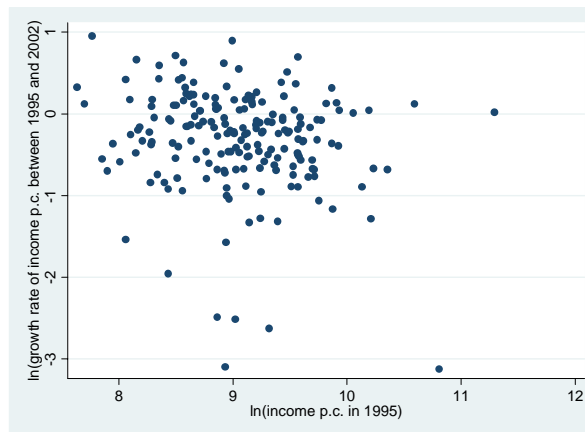
## Appendix B



Correlation between income p.c. and market access (1995)



Correlation between income p.c. and the spatial lag (1995)



Correlation between income p.c. in 1995 and  
its growth rate between 1995 and 2002

## Appendix C: Trade data

Trade equation estimations are carried out using trade flows from different sources to cover (i) intra-provincial (or intra-national), (ii) inter-provincial and (iii) international flows. Chinese and international trade flows are all merged into one single data set which allows us to calculate the market capacities of provinces and foreign countries based on their exports to all destinations (both domestic and international).

### C.1. International Data

International trade flows are expressed in current USD and come from IMF Direction of Trade Statistics (DOTS).

Intra-national trade flows are expressed in current USD and are calculated as the difference between domestic primary- and secondary-sector production minus exports. Production data for OECD countries come from the OECD STAN database. For other countries, the ratios of industrial and agricultural output as a percentage of GDP are extracted from Datastream. These are then multiplied by country GDP (in current USD) from World Development Indicators 2005.

### C.2 Chinese Data

Provincial foreign trade data are obtained from the Customs General Administration database, which records the value of all import and export transactions which pass via Customs. Provincial imports and exports are decomposed into those concerning up to 230 international partners. This database has previously been discussed by Lin (2005) and Feenstra, Hai, Woo and Yao (1998).

The exchange rate is the average exchange rate of the Yuan against the US dollar in the China Exchange Market. This comes from the China Statistical Yearbook.

Intra-provincial flows or foreign intra-national flows, i.e. exports to itself, are computed following Wei (1996) as domestic production minus exports. Production data for Chinese provinces are calculated as the sum of industrial and agricultural output. Output in yuan are converted into current USD using the annual exchange rate. All statistics come from China Statistical Yearbooks.

Inter-provincial trade is computed as trade flows with the rest of China. Provincial input-output tables<sup>21</sup> provide the decomposition of provincial output, and the international and domestic trade of tradable goods. These are available for 28 provinces, with data missing for Tibet, Hainan and Chongqing.

---

<sup>21</sup>Most Chinese provinces produced square input-output tables for 1997. A few of these are published in provincial statistical yearbooks. We obtained access to the final-demand columns of these matrices from the input-output division of China's National Bureau of Statistics. Our estimations assume that the share of domestic trade flows (that is between each province and the rest of China) in total provincial trade is constant over time.