
ELSEVIER
Editors
DANIEL McMILLEN
YVES ZENOU
Book Review Editor
ARTHUR O'SULLIVAN
Associate Editors
S.P. ANDERSON
L. ANSELIN
R. ARNOTT
M. BERLIANT
R. BOADWAY
J. BRUECKNER
A. CICCONE
P.-P. COMBES
N.E. COULSON
R. FORSLID
M. FUJITA
S. GABRIEL
E. GLAESER
J.V. HENDERSON
K. IHLANFELDT
R.P. INMAN
Y. IOANNIDES
M.E. KAHN
H. KONISHI
K.A. KONRAD
P. MARTIN
T. MAYER
T. MORI
T. NECHYBA
G. NORMAN
G.I.P. OTTAVIANO
H.G. OVERMAN
E. PATACCHINI
G. PERI
J.M. QUIGLEY
S. RAPHAEL
F. ROBERT-NICOUD
S.L. ROSS
E. SHIOJI
K.A. SMALL
K.O. STAHL
J.-F. THISSE
P. WANG
B. WEINBERG
J.E. ZABEL
Volume 41, Issue 1, January 2011 ISSN 0166 0462
regional SCIENCE & urban ECONOMICS
CONTENTS
J.N. van Ommeren and E. Gutiérrez-i-Puigarnau, Are workers with a long commute less productive? An empirical analysis of absenteeism 1
J. de Sousa and S. Poncet, How are wages set in Beijing? 9
A. Minniti and C.P. Parello, Trade integration and regional disparity in a model of scale-invariant growth 20
G.M. Ahlfeldt, Blessing or curse? Appreciation, amenities and resistance to urban renewal 32
G. Burge and C. Rogers, Local option sales taxes and consumer spending patterns: Fiscal interdependence under multi-tiered local taxation 46
C. Gaigné and I. Wooton, The gains from preferential tax regimes reconsidered 59
E. Anenberg, Loss aversion, equity constraints and seller behavior in the real estate market 67
Available online at www.sciencedirect.com
ScienceDirect

Regional Science and Urban Economics 41 (2011) 9–19

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

How are wages set in Beijing?

José de Sousa [a], Sandra Poncet [b,*]

[a] Centre d'Economie de la Sorbonne, Université de Paris 1, and CREM Université Rennes 1, France
[b] Paris School of Economics – Université de Paris 1 Panthéon-Sorbonne, and CEPII, France

ARTICLE INFO

Article history:
Received 11 February 2009
Received in revised form 19 July 2010
Accepted 21 July 2010
Available online 5 August 2010

JEL classification:
F12
F15
R11
R12

Keywords:
Wages
China
Migration
Economic geography

ABSTRACT

China's export performance over the past fifteen years has been phenomenal. Is this performance going to last? Wages are rising rapidly but a population in excess of one billion represents a large reservoir of labor. Firms in export-intensive provinces may draw on this reservoir to increase competition in their labor market and keep wages low for many years to come. We develop a wage equation from a New Economic Geography model to capture the upward pressure from national and international demand and downward pressure from migration. Using panel data at the province level, we find that migration has moderately slowed down Chinese wage increase over the period 1995–2007.

1. Introduction

China's export performance over the past fifteen years has been phenomenal. Its share of world merchandise exports jumped from 2.5% in 1993 to 5.9% in 2003 and 8.9% in 2007 (WTO, 2008). Imports have also grown but China's trade balance is substantially positive. This imbalance is a matter of concern for its main trade partners and the latest figures may aggravate the growing discontent. According to the EU trade commissioner Peter Mandelson, "the EU's trade deficit with China is growing $20 million an hour," (June 12 2007, Wall Street Journal).

Low wages are one of the main reasons for Chinese success in capturing world export markets. In 2004, Chinese average hourly manufacturing compensation is only U.S.$0.67, i.e. about 3% of the American average (Lett and Banister, 2006).[1] However, some analysts assert that this advantage is only temporary since wages are rising rapidly (Adams et al., 2006; Lett and Banister, 2006).[2] This upward wage trend may erode the once unbeatable China price and the competitiveness of the export-intensive provinces. On the other hand, we assert that a population in excess of one billion represents a large reservoir of labor. If labor is mobile across provinces, firms in export-intensive provinces may draw on this reservoir to increase competition in their labor market and keep wages low for many years to come. This paper attempts to shed some empirical light on this debate over wage setting in China.

Using a new economic geography (NEG) model, we estimate the maximum wage a firm in a given province can afford to pay given its provincial *market access* and *immigrant labor supply*. The provincial market access (or market potential) is defined as the world and internal demand each province faces given its geographical position and that of its trading partners (Harris, 1954; Redding and Venables, 2004). Wages are predicted to be higher in provinces with high levels of demand. There is some evidence to support this prediction in China (Lin, 2005; Hering and Poncet, 2010). Wages in coastal provinces with good market access, such as Fujian, Guangdong and Shanghai, are twice as high as the national average wage. The provincial immigrant

* Corresponding author.
E-mail addresses: jose.desousa@univ-paris1.fr (J. de Sousa), sandra.poncet@cepii.fr (S. Poncet).

[1] It is worth noting that low wages do not imply low production costs if productivity is also low. However, using the Penn World Tables 6.3 and a crude measure of productivity, we find that the Chinese output per worker represents around 12% of the American level. Thus, the US–Chinese difference in terms of wages appears to be larger than in terms of productivity. Another reason explaining the Chinese success in export markets is that its GDP accounts for an increasing share of world GDP. Thus, consistent with trade theory, China captures an increasing share of world merchandise exports.

[2] Chinese wages increased by 15% per year in 2001 and 2002 (Adams et al., 2006). Besides, total hourly compensation costs of manufacturing employees increased by nearly 18% between 2002 and 2004 (Lett and Banister, 2006). See also *China's competitiveness 'on the decline'*, Financial Times, March 22, 2006. Using our data set over 1995–2007 we confirm this upward trend (see Appendix A).

0166-0462/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2010.07.004

labor supply is defined as the supply of immigrant labor each province faces due to labor migration. An increase in the immigrant labor supply is expected to exert a downward pressure on wages.

Migration has long been severely restricted by a specific Chinese institution: the *hukou* system. The hukou is a system of household registration. It forces people to live and work in the place where they have an official registration. This system controls population movement and renders migration costly since provinces can impose various hurdles to obtaining the necessary registration (Au and Henderson, 2006). However, in the mid-1990s, the Chinese government has carried out a "deep reform" of the hukou system, notably to reduce restrictions on mobility (Wang 2004).[3] As a consequence of the hukou reforms, migration has increased dramatically. As an example, between 1995 and 2000, 25.78 million people between ages 15 and 64 moved from one province to another (Lin et al., 2004) and predominantly to the export-intensive Southern coastal provinces (Naughton, 2007).[4]

Using a data set covering 29 Chinese provinces between 1995 and 2007, we investigate the relative impact of provincial market access and immigrant labor supply on the average provincial manufacturing wage. We moreover control for the potential endogeneity of migration and other explanatory variables via instrumental variables and Generalized Method of Moments (GMM) estimators. The estimates suggest that provincial nominal wages increase by about 13.5% per year, as documented in the literature (Adams et al., 2006; Lett and Banister, 2006). However, we also find that migration has moderately slowed down Chinese wage increase over time. For instance, in the coastal province of Guangdong, making up approximately 12% of 2007 national GDP, immigration has slowed down wage growth by about 8.64% in total or 0.7% per year.

The paper is organized as follows. In the next section we outline the theoretical framework from which the econometric specification used in the subsequent sections is derived. In Section 3, we describe the data sources and discuss some estimation issues. In Section 4, we investigate econometrically the respective contributions of market access and immigrant share to the determination of wages in China. In Section 5, we conclude and discuss some implications of our results.

2. Theoretical framework

The theoretical framework underlying the empirical analysis is based on a standard new economic geography model (Fujita et al., 1999; Redding and Venables, 2004). We add worker skill heterogeneity to control for differences in productivity across regions. Such differences explain part of the spatial wage disparities (Combes et al., 2008). Using this simple model, we first derive a so-called short-run equilibrium (see Krugman, 1991), taking as given the allocation of workers in each region. Then, we propose a simple strategy, inspired by the labor economics literature, to estimate the short-run impact of a labor immigration shock on wages.

2.1. A short-run equilibrium

The economy is composed of $i=1,\ldots,R$ regions. Each region produces differentiated goods under increasing returns. The utility of a representative agent in region j has a standard Dixit–Stiglitz form:

$$U_j = \left[\sum_{i=1}^{R} n_i q_{ij}^{(\sigma-1)/\sigma}\right]^{\sigma/(\sigma-1)}, \quad \sigma > 1, \tag{1}$$

where q_{ij} denotes the demand in region j for a variety produced in region i, n_i is the number of varieties produced in region i, and $\sigma > 1$ is the elasticity of substitution. Dual to this quantity aggregator is the price index for each region:

$$G_j = \left[\sum_{i=1}^{R} n_i \left(p_i T_{ij}\right)^{1-\sigma}\right]^{1/(1-\sigma)}, \tag{2}$$

where p_i is the mill price of products made in i and $T_{ij} > 1$ is the usual iceberg trade costs between region i and region j. Transporting products from one region to another is costly. Thus, the iceberg transport technology assumes that for every unit of good shipped, only a fraction $\left(\frac{1}{T_{ij}}\right)$ arrives.

As is well known the value of demand for a variety produced in i and sold in j is given by

$$q_{ij} = \left(p_i T_{ij}\right)^{-\sigma} G_j^{\sigma-1} E_j, \tag{3}$$

where E_j is total expenditure of region j. 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:

$$q_i = \sum_{j=1}^{R} \left(p_i T_{ij}\right)^{-\sigma} G_j^{\sigma-1} E_j T_{ij} = p_i^{-\sigma} MA_i, \tag{4}$$

where

$$MA_i = \sum_{j=1}^{R} T_{ij}^{1-\sigma} G_j^{\sigma-1} E_j, \tag{5}$$

is the market access of exporting region i (Redding and Venables, 2004: 59). This is given by a trade cost (T_{ij}) and price index (G_j) weighted sum of the regional expenditures (E_j).

In this simple model, labor is the only factor of production. Following Head and Mayer (2006), we introduce skill heterogeneity and assume that the labor requirement, ℓ, depends on both output per firm, q, and average years of schooling, h, as follows:

$$\ell_i = (F + cq_i) exp(-\rho h_i), \tag{6}$$

where F and c represent fixed and marginal requirements in "effective" (education-adjusted) labor units. The parameter ρ measures the return to education and shows the percentage increase in productivity due to an increase in the average population share with higher education. Maximizing profits yields the familiar mark-up pricing rule for varieties produced in region i:

$$p_i = \frac{\sigma}{\sigma-1} w_i\, c\, exp(-\rho h_i), \tag{7}$$

Free entry implies that the equilibrium output of any firm is:

$$q^* = \frac{F(\sigma-1)}{c}. \tag{8}$$

Using the demand function (4), the pricing rule (7) and equilibrium output (8), we get the manufacturing wage when firms break even:

$$w_i = \frac{\sigma-1}{\sigma\, c\, exp(-\rho h_i)} \left[MA_i \frac{c}{F(\sigma-1)}\right]^{1/\sigma}. \tag{9}$$

Despite the lack of any explicit dynamics in the model, it is useful to consider wage Eq. (9) as a short-run equilibrium, taking as given the allocation of workers in each region. This equilibrium is consistent with the existence of frictions in labor mobility across Chinese

[3] Barriers to migration have been relaxed since the mid 1980s (Poncet, 2006) but remained tight (see Au and Henderson, 2006).

[4] Lin et al. (2004) document "a significant redirection of migration flows in 1995–2000." Over this period inland-to-coast migration nearly doubled to reach a 60.1% share of all inter-provincial migration.

provinces due to the hukou system of registration (see Au and Henderson, 2006).

2.2. Deviation from the short-run equilibrium

If restrictions on mobility remain tight, reforms of the hukou system have lowered the barriers to migration and have increased migration flows (see previous discussion). Based on this evidence, we now work out a deviation from the short-run equilibrium by considering the effect of an influx of immigrants in a given region i on wages w_i. To this end, we proceed in several steps. First, we rearrange the wage equilibrium Eq. (9) by determining the equilibrium labor demand for workers in region i. From Eq. (6), we can solve for q using the equilibrium output solution determined in Eq. (8) and get

$$\ell_i^* = \sigma F\, exp(-\rho h_i). \tag{10}$$

Then, we turn Eq. (10) around and express fixed requirements as:

$$F = \frac{\ell_i^*}{\sigma\, exp(-\rho h_i)}. \tag{11}$$

Replacing Eq. (11) in the wage equilibrium Eq. (9) gives:

$$w_i = (\sigma-1)^{\frac{\sigma-1}{\sigma}}(MA_i)^{\frac{1}{\sigma}}[c\sigma\, exp(-\rho h_i)]^{\frac{1-\sigma}{\sigma}}\ell_i^{*-\frac{1}{\sigma}}. \tag{12}$$

Second, we benefit from the multiplicative form of Eq. (12) to operate a log-linear transformation of the model. In addition, we introduce time subscripts $t=1,...,T$ to stress the point of a deviation from the short-run equilibrium

$$ln\, w_{it} = \alpha_0 + \alpha_1 ln\, MA_{it} + \alpha_2 h_{it} + \alpha_3 ln(\ell_{it}^*), \tag{13}$$

where α_0 gathers the constant terms, $\alpha_1 = \frac{1}{\sigma}$, $\alpha_2 = \frac{\sigma-1}{\sigma}\rho$, and $\alpha_3 = -\frac{1}{\sigma}$, i indexes region and t time.

Finally, we depart from the short-run equilibrium situation (12) by following the methodology of Friedberg (2001) and Borjas (2003) in labor economics. Friedberg (2001) and Borjas (2003) estimate the effect of an immigrant penetration on wages by considering an exogenous influx of immigrant m_i in the region of destination i.[5] The rate of change in labor due to the immigration is defined as a simple first-difference $ln(\ell_{it+1}) - ln(\ell_{it}) = ln(\ell_{it} + m_{it}) - ln(\ell_{it}) = ln\left(\frac{\ell_{it}+m_{it}}{\ell_{it}}\right)$. Letting $z \equiv \left(\frac{\ell_{it}+m_{it}}{\ell_{it}}\right)$ and noting that $lim_{z\to 1} lnz = z-1$, we deduce that $ln\left(\frac{\ell_{it}+m_{it}}{\ell_{it}}\right) \approx \left(\frac{\ell_{it}+m_{it}}{\ell_{it}}\right) - 1 = \frac{m_{it}}{\ell_{it}}$. Consequently, the wage change resulting from an exogenous influx of immigrants is given by

$$\Delta ln\, w_{it} = \alpha_1 \Delta ln\, MA_{it} + \alpha_2 \Delta h_{it} + \alpha_3 \frac{m_{it}}{\ell_{it}}, \tag{14}$$

where the Δ prefix denotes the change from one time period to the next and $\Delta ln(\ell_{it}) \approx \frac{m_{it}}{\ell_{it}}$ is the immigrant share of region i. The predictions of Eq. (14) are that (1) the regression coefficients $\alpha_1 = \frac{1}{\sigma}$ and $\alpha_2 = \frac{\sigma-1}{\sigma}\rho$ are positive; (2) the regression coefficient $\alpha_3 = -\frac{1}{\sigma}$ is negative; (3) the estimated coefficients α_1 and α_3 are equal in absolute value, and (4) the elasticity of substitution (σ) between traded goods is greater than one.

[5] The exogeneity assumption of the immigrant influx is supported by analytical work corroborating that the hukou system reforms have dramatically increased migration in China (see earlier discussion). However, we cannot entirely rule out the possibility that differential of wages across regions could have stimulated migration. We investigate empirically this endogeneity issue.

3. Data and endogeneity issues

The core empirical part of this paper explains the variation in average provincial manufacturing wages in China. Before turning to the endogeneity issues of the explanatory variables, we first describe the data.

3.1. Data

The data set covers 29 Chinese provinces over the period 1995–2007.[6] Thus, we get 377 ($=29*13$) observations. Our explained variable is the average annual nominal wage rate of manufacturing workers and staff in a province. This is defined as the ratio of the total wage bill to the number of manufacturing workers and staff by province and year.

We detail now the construction of the market access and the immigrant share variables. Appendix B provides details and sources about the other explanatory variables and Appendix C provides summary statistics for all of the variables.

3.1.1. Construction of market access

The market access variable is defined as a trade cost and price index weighted sum of the provincial expenditures (Eq. 5). In order to compute the market access variable, we exploit information from the estimation of bilateral trade as in Redding and Venables (2004). Summing Eq. (3) over all of the products produced in region i, we obtain the total value of the exports of i to j:

$$X_{ij} = n_i\left(p_i T_{ij}\right)^{1-\sigma} G_j^{\sigma-1} E_j = sc_i \phi_{ij} mc_j, \tag{15}$$

where $sc_i = n_i(p_i)^{1-\sigma}$ measures the "supply capacity" of the exporting region i, $mc_j = G_j^{\sigma-1}E_j$ measures the "market capacity" of the importing region j, and $\varphi_{ij} = T_{ij}^{1-\sigma}$ denotes the "freeness" of trade (Baldwin et al., 2003).[7] In Redding and Venables (2004), trade costs only depend on bilateral distance between regions. In our empirical approach, regions refer to Chinese provinces as well as international countries. Consequently, we allow for differentiated trade cost measures depending on whether trade occurs within province, within country, between provinces,[8] between countries or between provinces and countries. Thus freeness of trade depends on bilateral distance ($dist_{ij}$) but also on a set of dummies indicating whether provincial or foreign borders are crossed.

$$\phi_{ij} = dist_{ij}^{-\delta} exp\left[-\varphi B_{ij}^f - \varphi^* B_{ij}^{f^*} + \psi Contig_{ij} - \vartheta B_{ij}^c + \xi B_{ij}^i + \zeta_{ij}\right], \tag{16}$$

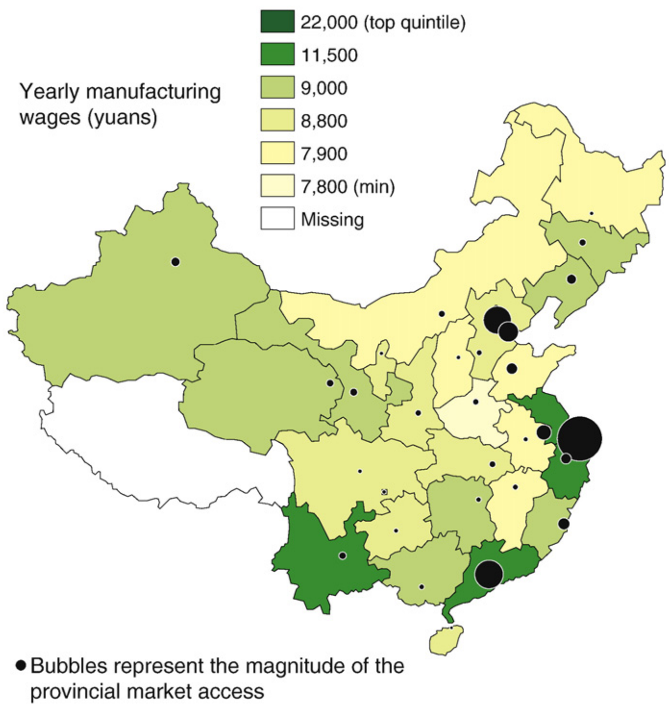
where $B_{ij}^f = 1$ if i is a Chinese province and j a foreign country or i a foreign country and j a Chinese province; $B_{ij}^{f^*} = 1$ if $i \neq j$ are two foreign countries; $Contig_{ij} = 1$ if $i \neq j$ are two foreign countries sharing a land border; $B_{ij}^c = 1$ if $i \neq j$ are two Chinese provinces[9]; $B_{ij}^i = 1$ if $i = j$ denote the same foreign country and 0 otherwise. The error ζ_{ij}

[6] China is divided into 27 provinces plus four province-status "super-cities" — Beijing, Chongqing, Shanghai and Tianjin. We cover all of the provinces apart from Tibet due to missing observations. Moreover, Chongqing and Sichuan are considered together.

[7] $\varphi_{ij} \in [0,1]$ equals 1 when trade is free and 0 when trade is eliminated due to high shipping costs and elasticity of substitution (σ).

[8] Note that we face severe data constraints to measure inter-provincial trade in China, since data on bilateral trade between Chinese provinces do not exist. As a consequence, we use provincial input–output tables to proxy inter-provincial trade of a province as its trade with the rest of China. Refer to Appendix D1 for details.

[9] Since inter-provincial trade in China is measured by trade between provinces and the rest of China (see footnote 8), contiguity between Chinese regions is entirely redundant with the B_{ij}^c dummy.



Map 1. Nominal manufacturing wages and constructed market access in China (2002).

captures the unmeasured determinants of trade freeness. Consequently, this specification allows the impediments to domestic trade to be different from the impediments to international trade (see Appendix D for details).

We specify the following log-linear trade regression by first substituting Eq. (16) into Eq. (15), and then capturing unobserved exporting ($lnsc_i$) and importing ($lnmc_j$) region characteristics with exporter and importer fixed effects (cty_i and ptn_j):

$$lnX_{ij} = ln\mu + cty_i + ptn_j - \delta \, ln \, dist_{ij} - \varphi B^f_{ij} - \varphi^* B^{f*}_{ij} + \psi Contig_{ij} - \vartheta B^c_{ij} + \xi B^i_{ij} + \zeta_{ij}. \quad (17)$$

Using a complete data set of trade (see Appendix D),[10] we estimate Eq. (17) in cross-section, for each year of our sample. The yearly estimated coefficients are then used to construct predicted values for market access defined as $MA_i = \sum_{j=1}^{R} \varphi_{ij} mc_j$.[11] Our estimated market access variable consists of three parts: local (intra-provincial demand), national (demand from the rest of China), and worldwide:

$$\begin{aligned}\widehat{MA}_i &= \hat{\phi}_{ii} G_i^{\sigma-1} E_i + \sum_{j \in P} \hat{\phi}_{ij} G_j^{\sigma-1} E_j + \sum_{j \in F} \hat{\phi}_{ij} G_j^{\sigma-1} E_j \\ &= exp(ptn_i) \times dist_{ii}^{-\hat{\delta}} + \sum_{j \in P} exp\left(ptn_j\right) \times dist_{ij}^{-\hat{\delta}} \times exp(\vartheta) \quad (18) \\ &\quad + \sum_{j \in F} exp\left(ptn_j\right)^{\hat{\lambda}} \times dist_{ij}^{-\hat{\delta}} \times exp\left(\hat{\varphi} + \hat{\psi} Contig_{ij}\right),\end{aligned}$$

where P and F stand for Chinese provinces and foreign countries, respectively. The results for various years are presented in Table 3 in Appendix D3. Map 1 shows that coastal provinces had greater market access and higher nominal wages in manufacturing in 2002. This pattern is confirmed in the various figures in Appendix D4.

3.1.2. Immigrant share

We rely on the annual Sample Survey on Population Changes to compute $(\frac{m_i}{\ell_i})$ as the number of newly arrived non-residents divided by the population.[12] We actually assume that the number of newly arrived non-residents in a province is a good proxy for the immigrant labor supply (m_i). Newly arrived non-residents in a province are defined as the population living in "township, towns and street communities with permanent household registration elsewhere, [and] having been away from that place for less than six months" as well as the population "with residence registration in this enumeration area not yet settled". As a robustness check, we will verify that our results remain unchanged when this later category ("with residence registration in this enumeration area not yet settled") is excluded from the migration variable construction.

3.2. Endogeneity issues

When estimating Eq. (14), we face a potential endogeneity issue related to our explanatory variable of interest, migration. The literature documents that political reforms of the hukou system have reduced frictions in mobility and dramatically increased migration (Wang 2004). However, we cannot rule out the possibility that migration decisions are also driven by an income differential between the origin and destination locations. Consequently, the exogeneity assumption of the labor supply shock may be a concern. To deal with this issue we use initially an instrumental variable approach. The reliability of this method lies on the identification of instruments which are correlated with migration but uncorrelated with the error term, i.e. with the unobserved component of wages. An exogenous source of variation in migration may be found in climate variables (see Roback, 1982). We argue that unfavorable climate conditions in the origin location of the migrant may increase its probability of migration. We consider two complementary climate dimensions: annual *temperature* and annual *rain* (precipitation) in major cities of a province. Both dimensions may affect migration. Lin et al. (2004) and Bao et al. (2007) document that temperatures are significant in explaining migration in China: "People seem to move to warm areas" (Lin et al., 2004). In addition, there is evidence that water scarcity and drought in China cause internal migration. In the beginning of the 1990s, Homer-Dixon (1994) documents that "tens of millions of Chinese are trying to migrate from the country's interior and northern regions, where water and fuelwood are desperately scarce and the land is often badly damaged, to the booming coastal cities." We thus instrument immigration in destination province i based on yearly information on each climate condition in origin provinces j as an annual weighted average:

$$I_{temp^t_i} = \sum_{j \neq i} temperature^t_j \frac{immigration_{ij}}{\sum_{j \neq i} immigration_{ij}},$$

$$I_{rain^t_i} = \sum_{j \neq i} rain^t_j \frac{immigration_{ij}}{\sum_{j \neq i} immigration_{ij}},$$

where the weight is the share of the province of origin j in the total immigration of province i (termed "$immigration_{ij}$") between 1985 and

[10] A complete data set of trade is required to estimate appropriately the market capacity of each trade partner (within China or abroad).

[11] Eq. (17) allows also to construct a supplier access variable, $SA_j (= \sum_i \varphi_{ij} sc_i)$. However, since market access and supplier access variables tend to be highly correlated (see Amiti and Javorcik (2008) in the Chinese context), we follow most of the NEG literature and concentrate on market access forces.

[12] The results remain unchanged if we use the number of permanent residents in a province as a proxy for ℓ_i; these are available upon request. Permanent residents are defined as the population "residing in township, towns and street communities with permanent household registration there", i.e. in province i.

1990.[13] To reinforce the exogeneity of these two instruments, we exclude information about the destination province i in their calculation. We also introduce in the wage equation (i.e. the second stage) the annual averages of temperature and precipitation of the major cities in province i as additional control variables. This ensures that the instruments are not simply proxying the climate conditions in the destination province i.

We are however concerned about the potential endogeneity of other explanatory variables (such as education and market access). Thus, we also estimate the wage equation using a generalized method of moments (GMM) estimator. Additionally, we will verify that our results remain unchanged when our two external climate instruments are added to the internal instruments in the GMM estimations.

4. Estimation results

We now proceed to the estimation of our reduced-form wage equation on a panel of 29 provinces over the period 1995–2007.

4.1. Benchmark estimates

Table 1 reports the results of three different estimators: the two-stage least squares (2SLS), the limited information maximum likelihood (LIML) and the Generalized Method of Moments (GMM). While the first two rely on external instruments and address the endogeneity of our main variable of interest (migration), the third resorts to internal instruments to address the potential endogeneity of all the explanatory variables (including market access and education).

In the first four columns, we use an instrumental variable approach combined with provincial fixed effects to address the potential endogeneity of migration. The use of provincial fixed effects is an alternative to the first differences. It gives similar estimates than differencing and avoids losing the first time period for each cross-section (see Wooldridge, 2002). We have T time periods for each i, rather than $T-1$. The provincial fixed effects capture province-specific (time-independent) factors such as their surface areas, maritime access, distance to major partners and landlockness. These (unobserved) factors are important to capture since "cross-region variation in worker characteristics may reflect regional characteristics that are constant over the sample period" (Hanson, 2005). In addition, we introduce year fixed effects to capture the unobserved effects of economy-wide changes over time. We also cluster standard errors at the province level. Finally, we follow Borjas (2003) and address the interpretation problem that a rise in the immigrant share can represent either an increase in the number of non-residents or a fall in population. We thus add the province's population level in all the specifications. Our wage equation fits quite well the data by explaining more than 90% of the within variation in provincial wages.[14]

In columns 1 and 2, we instrument the immigrant share solely. Relying on an instrumental variable approach allows us to control for any simultaneity between wages and migration. As described earlier, we appeal to two different instruments (I_{temp} and I_{rain}), based on the climate conditions of the original locations. To ensure that our instruments are not simply proxying the climate conditions in the destination province i, we also introduce in the wage equation the annual averages of temperature and precipitation of the major cities in province i as additional control variables. As a precondition for the reliability of the instrumental variables procedure, we check the validity of our instruments via the Hansen test of overidentifying restrictions. The resulting insignificant test statistics, reported at the bottom of Table 1, indicate that the orthogonality of the instruments to the error term cannot be rejected, suggesting that our instruments are appropriate.

The first-stage, common to the 2SLS and LIML estimations, is reproduced in column 1 of Table 4 in Appendix E. We find that both our climate based instruments enter significantly and with the expected (negative) signs. First-stage results suggest that warmer temperature and larger rainfall in the origin locations reduce the intensity of emigrating to the destination province. However, the F-statistic on the excluded instruments, reported at the bottom of Table 1, is slightly below 10, the informal threshold suggested by Staiger and Stock (1997) to assess the validity of instruments. This raises a concern of weak instruments. To address this issue we first report the Stock and Wright (2000) statistic that provides weak-instrument robust inference for testing the significance of the endogenous regressors. We do not reject the null hypothesis (with the p-value of 0.12) that the coefficients of the excluded instruments are jointly equal to zero. In addition, we check the 2SLS estimates of column 1 with LIML in column 2. LIML is less precise but also less biased (Angrist and Pischke, 2009). Yet, the LIML estimates look fairly similar to 2SLS.

On the plus side, the LIML standard errors are not too far above the 2SLS standard errors. Together the Stock and Wright test and the similarity of LIML and 2SLS estimates tend to support the validity of our excluded instruments.[15]

In Table 1, from column 3 onward, we also instrument the population level with deep lags (15 years). In columns 3 and 4, partial F-statistics are now higher and slightly above 10 for population, indicating that the excluded instruments provide a reasonably good fit in the first-stage regressions. We report the first-stage of the additional instrumented variable (i.e. population) in column 2 of Table 4. We find that the 15-year lagged population has a significant positive impact on current population. The Sargan overidentification tests (with high p-values) support the validity of the instruments. Again, it is reassuring that LIML estimates (column 4) are almost identical to 2SLS (column 3). These results are fairly similar to those of columns 1 and 2, with the exception of the population estimate, which is now statistically insignificant.

We verify that our results are robust when relying on longer time intervals. So far our estimates are obtained using thirteen yearly intervals for each province. We use instead two different sets of sub-periods: a first set of four intervals of three-year (1995, 1998, 2001, and 2004) and a second set of three intervals of four-year (1995, 2001 and 2007). The results are reported in Table 5 in Appendix E. The three columns use the same specification as column 1 in Table 1. The sample size differs across columns: Column 1 uses the full sample, column 2 the years 1995, 1998, 2001, 2004 and 2007, and column 3 the years 1995, 2001 and 2007. The number of observations decreases with the length of intervals from 377 to 145 and 87. However, our main findings (a negative and significant impact of immigration shock and a positive and significant impact of market access) remain even if weaker.

In columns 5 and 6 of Table 1 we rely on internal instruments, using the GMM panel estimator proposed by Arellano and Bond (1991).[16] Eq. (14) is estimated in first-difference. As a result, the number of observations drops from 377 to 348. The instruments for the regression in differences are the right-hand side variables expressed

[13] Such bilateral measures of migration are built from census data (National Bureau of Statistics of China, 2002) and are only available every five years.

[14] Since the predicted values of market access are generated from a previous trade regression, we check the sensitivity of our results using bootstrapped standard errors: the results remained unchanged. The bootstrapped standard errors (500 replications) are available upon request.

[15] Our findings are coherent with Angrist and Pischke (2009): 215) who claim that "This suggests that you can't always determine instrument relevance using a mechanical rule, such as 'F>10'".

[16] We thank an anonymous referee for this suggestion.

Table 1
Manufacturing wage equation.

	Dependent variable: *ln*(manufacturing wage)					
Instrumented variables	MIG		MIG, POP		All variables	
Model	(1): 2SLS	(2): LIML	(3): 2SLS	(4): LIML	(5): GMM	(6): GMM augmented
Immigrant share	-0.132^b	-0.135^b	-0.140^b	-0.141^b	-0.024^b	-0.024^b
	(0.067)	(0.069)	(0.062)	(0.063)	(0.011)	(0.011)
ln(market access)	0.050^b	0.050^b	0.047^c	0.047^c	0.106^a	0.106^a
	(0.021)	(0.021)	(0.025)	(0.025)	(0.037)	(0.037)
Higher-education ratio	0.747^b	0.761^b	0.690	0.701	0.366^a	0.366^a
	(0.367)	(0.377)	(0.476)	(0.484)	(0.132)	(0.131)
ln(population)	-0.582^b	-0.577^b	−0.342	−0.348	-1.150^a	-1.150^a
	(0.237)	(0.240)	(0.572)	(0.577)	(0.352)	(0.354)
IN(rain)	0.012	0.012	0.012	0.012	0.032^b	0.032^b
	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)	(0.015)
Temperature	0.002	0.002	0.002	0.002	0.00006	0.000003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.008)	(0.008)
Year 1996[h]	0.112^a	0.112^a	0.109^a	0.110^a	0.118^a	0.118^a
Year 1997	0.144^a	0.144^a	0.139^a	0.139^a	0.180^a	0.181^a
Year 1998	0.308^a	0.307^a	0.300^a	0.300^a	0.370^a	0.370^a
Year 1999	0.379^a	0.378^a	0.370^a	0.370^a	0.462^a	0.463^a
Year 2000	0.543^a	0.542^a	0.539^a	0.539^a	0.599^a	0.600^a
Year 2001	0.642^a	0.641^a	0.636^a	0.636^a	0.719^a	0.720^a
Year 2002	0.745^a	0.744^a	0.736^a	0.736^a	0.849^a	0.850^a
Year 2003	0.854^a	0.852^a	0.845^a	0.844^a	0.961^a	0.962^a
Year 2004	0.991^a	0.989^a	0.981^a	0.980^a	1.111^a	1.112^a
Year 2005	0.992^a	0.990^a	0.984^a	0.983^a	1.112^a	1.112^a
Year 2006	1.241^a	1.239^a	1.229^a	1.228^a	1.385^a	1.385^a
Year 2007	1.353^a	1.350^a	1.338^a	1.337^a	1.525^a	1.525^a
Observations	377	377	377	377	348	348
Within R^2	0.978	0.977	0.976	0.976		
Test[i] H0: $\alpha_1+\alpha_3=0$	$\chi^2(1)=1.5$	$\chi^2(1)=1.5$	$\chi^2(1)=2.0$	$\chi^2(1)=2.0$	F(1,29) = 3.9	F(1,29) = 3.9
First-stage						
Migration R^2 (%)	2.7	5.88				
Migration F-statistic	6.2^a	4.85^b				
ln(population) R^2 (%)			13.33			
ln(population) F-stat			6.53^a			
Weak instr. test[j] χ^2	4.21	3.30				
P-value	0.12	0.35				
Overid. test[k] χ^2	0.32	0.31	0.28	0.28	11.11	17.75
AR(1)[l]					-2.79^a	-2.82^a
AR(2)					−0.59	−0.59

Notes: The table compares 2SLS, LIML and GMM estimates. First-differencing in the GMM estimates results in the drop of one year (sample size goes from 377 to 348). Clustered standard errors at the province level are reported in parentheses. [a,b] and [c] denote significance at the 1%, 5% and 10% levels respectively. Immigrant share is defined as non-residents over population. Higher-education ratio is measured as the share of above 6 years old population with at least high school education. [h]To save space, we do not report the constant and the clustered standard errors of the year dummies (they are available upon request). Depending on the specification in columns 1 to 4, instrumented variables are *immigrant share* and *population*; and excluded instruments are *climate* variables (I_{temp} and I_{rain}), and 15-year lagged *population*. See the text for more details and Appendix E for the first-stage results. [i] The tests in all columns fail to reject (at the 5% confidence level) the null hypothesis that the coefficients on immigrant share and market access are equal in absolute value. [j]Stock–Wright weak-instrument-robust inference test. [k]The Hansen test for overidentifying restrictions does not reject the null hypothesis that our instruments are appropriate. [l] Significant negative first-order (AR(1)) serial correlation in differenced residuals and no evidence of second-order serial correlation (AR(2)) in the first-differenced residuals indicate that the orthogonality conditions cannot be rejected at the one percent level, thus that the error term is not serially correlated.

in level, lagged twice or more. Consistency of the GMM estimates depends on the validity of the instruments. We check their validity using two tests suggested by Arellano and Bond (1991): the test for overidentifying restrictions and the test for second-order serial correlation of the residuals (AR(2)).[17] The AR(2) test is asymptotically distributed as a standard normal distribution under the null of no second-order serial correlation. It provides a further check on the specification of the model and on the legitimacy of variables dated t-2 as instruments. Overidentifying restrictions will be tested based on the Hansen test that is asymptotically distributed as a chi-square with degrees of freedom equal to the number of instruments less the number of parameters, under the null of instrument validity. Compared to column 5, column 6 adds (I_{temp} and I_{rain}) to the list of instruments. As can be seen, this modification leaves the results virtually unchanged. The tests of the consistency of the GMM estimates are reported at the bottom of the Table 1. The AR(2) test indicates that the orthogonality conditions cannot be rejected at the one percent level, thus that the error term is not serially correlated. The Hansen test for overidentifying restrictions does not reject the null hypothesis that our instruments are appropriate. Results based on the GMM estimator confirm that manufacturing wages in China rise with market access and decline in response to positive immigration shocks.

[17] We also report tests for first-order and second-order serial correlation in the first-differenced residuals. If the disturbances ε_{it} are not serially correlated, there should be evidence of a significant negative first-order serial correlation in differenced residuals and no evidence of second-order serial correlation in the first-differenced residuals. Significant second-order serial correlation of the first-differenced residuals indicates that the original error term is serially correlated and thus that the instruments are misspecified. Alternatively, if the test fails to reject the null hypothesis of no second-order serial correlation, we conclude that ε_{it} is serially uncorrelated and the moment conditions are well specified.

Table 2
Manufacturing wage equation.

Model	Dependent variable: *ln*(manufacturing wage) GMM estimator					
	(1)	(2)	(3)	(4)	(5)	(6)
	Benchmark	No cities	No unsettled registration	Female migration	Male migration	Primary education
Immigrant share	−0.024^{b}	−0.038^{b}	−0.019^{b}	−0.028^{b}	−0.021^{c}	−0.018^{c}
	(0.011)	(0.016)	(0.008)	(0.011)	(0.012)	(0.010)
ln(market access)	0.106^{a}	0.068^{c}	0.099^{a}	0.102^{a}	0.061^{b}	0.087^{b}
	(0.037)	(0.035)	(0.035)	(0.036)	(0.024)	(0.033)
Higher-education ratio	0.366^{a}	0.116	0.320^{b}	0.367^{a}	0.224^{c}	0.156
	(0.132)	(0.275)	(0.125)	(0.125)	(0.111)	(0.114)
ln(population)	−1.150^{a}	−1.445^{b}	−0.957^{a}	−1.161^{a}	−0.824^{a}	−0.876^{a}
	(0.352)	(0.531)	(0.278)	(0.354)	(0.200)	(0.288)
ln(rain)	0.032^{b}	0.049^{a}	0.027^{b}	0.030^{c}	0.022^{c}	0.027^{b}
	(0.016)	(0.017)	(0.011)	(0.015)	(0.011)	(0.011)
Temperature	0.000	0.010	0.000	0.000	0.006	0.002
	(0.008)	(0.009)	(0.006)	(0.009)	(0.008)	(0.006)
Year 1996^{h}	0.118^{a}	0.114^{a}	0.102^{a}	0.117^{a}	0.097^{a}	0.102^{a}
Year 1997	0.180^{a}	0.181^{a}	0.163^{a}	0.178^{a}	0.155^{a}	0.165^{a}
Year 1998	0.370^{a}	0.344^{a}	0.350^{a}	0.366^{a}	0.324^{a}	0.349^{a}
Year 1999	0.462^{a}	0.448^{a}	0.442^{a}	0.458^{a}	0.422^{a}	0.448^{a}
Year 2000	0.599^{a}	0.617^{a}	0.581^{a}	0.594^{a}	0.564^{a}	0.587^{a}
Year 2001	0.719^{a}	0.734^{a}	0.698^{a}	0.715^{a}	0.677^{a}	0.709^{a}
Year 2002	0.849^{a}	0.853^{a}	0.826^{a}	0.846^{a}	0.796^{a}	0.839^{a}
Year 2003	0.961^{a}	0.977^{a}	0.938^{a}	0.959^{a}	0.916^{a}	0.958^{a}
Year 2004	1.111^{a}	1.135^{a}	1.085^{a}	1.107^{a}	1.056^{a}	1.107^{a}
Year 2005	1.112^{a}	1.135^{a}	1.089^{a}	1.108^{a}	1.059^{a}	1.106^{a}
Year 2006	1.385^{a}	1.401^{a}	1.359^{a}	1.381^{a}	1.324^{a}	1.378^{a}
Year 2007	1.525^{a}	1.542^{a}	1.500^{a}	1.521^{a}	1.468^{a}	1.525^{a}
Observations	348	312	348	348	348	348
Test^{i} H0: $\alpha_1+\alpha_3=0$	F(1,29) = 3.9	F(1,26) = 0.40	F(1,29) = 4.7^{x}	F(1,29) = 3.4	F(1,29) = 2.3	F(1,29) = 3.5
Overid. test^{j} χ^2	11.11	9.89	11.23	13.91	10.33	12.06
AR(1)^{k}	−2.79^{a}	−3.20^{a}	−2.80^{a}	−2.78^{a}	−2.19^{b}	−2.59^{b}
AR(2)	−0.59	−0.62	−0.48	−0.53	−0.51	−0.32

Notes: The sample size is 348 except in column (2) where 36 observations for three province-level cities (Beijing, Tianjin and Shanghai) over 12 years are dropped from the sample. Clustered standard errors at the province level are reported in parentheses. ^{a}, ^{b} and ^{c} denote significance at the 1%, 5% and 10% levels respectively. Immigrant share is defined as non-residents over population in columns (1), (2), (3) and, (6); as female non-residents over population in column (4), and as male non-residents over population in column (5). In column (6) non-residents do not include population with unsettled registration. Higher-education ratio is measured as the share of above 6 years old population with at least high school education in all columns but column (6), where it is measured as the share of above 6 years old population with at least *primary* education. ^{h}To save space, we do not report the constant and the standard errors of the year dummies (they are available upon request). ^{i}The tests in all columns (but column 3 for which ^{x} indicates that H0 is rejected at the 5% confidence level) fail to reject (at the 5% confidence level) the null hypothesis that the coefficients on immigrant share and market access are equal in absolute value. ^{j}The Hansen test for overidentifying restrictions does not reject the null hypothesis that our instruments are appropriate. ^{k}Significant negative first-order (AR(1)) serial correlation in differenced residuals and no evidence of second-order serial correlation (AR(2)) in the first-differenced residuals indicate that the orthogonality conditions cannot be rejected at the one percent level, thus that the error term is not serially correlated.

While the flavor of the results is similar across the three estimation methods, the GMM estimator, which addresses the endogeneity of all the explanatory variables, suggests a much lower impact of higher education and immigration and a much higher impact of market access. Based on the GMM results of column 5 (our preferred specification), it is useful to interpret the magnitude of the estimated coefficients. Holding other factors constant, a 10% increase in market access raises wages by about 1.06% on average. In addition, a one-point increase in the share of population with at least secondary education raises wages roughly by 0.4%. Finally, a one-point increase in the immigrant ratio decreases wages by about 2.4%.

Before elaborating on these results and on the negative impact of the immigrant share estimate, let us first check the sensitivity of our results.

4.2. Robustness checks

Table 2 reports additional robustness checks obtained using the GMM estimator. The first column reproduces our benchmark, i.e. column 5 of Table 1. In column 2, we verify that our results are not driven by the specific features of the "super-cities".[18] We test whether the "municipality" provinces are outliers by dropping the observations for the three "super-cities" of our sample (Beijing, Tianjin and Shanghai).[19] The sample size drops from 348 to 312 observations but leaves our results almost unchanged. We conclude that these super-cities do not drive all the results and thus keep them in the following regressions.

In column 3, we verify the robustness of our results by using an alternative proxy for the migration share. We restrict the definition of the newly arrived non-residents in a province to the population living in "township, towns and street communities with permanent household registration elsewhere, [and] having been away from that place for less than six months". We thus exclude the population "with residence registration in this enumeration area not yet settled". This modification does not affect significantly our benchmark estimates.

In columns 4 and 5, we check for a gender effect. Current international migration is different from past mass migration, when migrants were disproportionately men (Freeman, 2006). As in current international migration, nearly half of the current migrants in China

[18] We thank an anonymous referee for this suggestion. Two of the "super-cities" stand out for their relatively large immigrant shares. The top provinces in terms of immigrant shares for 2007 are Guangdong (5.2%), Beijing (2.9%), Fujian (2.5%), Zhejiang (1.7%) and Shanghai (1.6%).

[19] Since Chongqing has been quite recently (in 1997) carved out of the Sichuan province, we choose, to solve data availability problems, to follow the empirical literature and consider Chongqing and Sichuan together and thus reconstitute the pre-1997 Sichuan province. As such our sample only includes three "super-cities": Beijing, Tianjin and Shanghai.

are women (see Appendix C). Our results still hold if we take this new trend into account and redefine the immigrant share as female non-residents over population in column (4), and as male non-residents over population in column (5). Again, our results are not significantly modified.

In column 6, we verify the robustness of our results by using an alternative proxy for the education level. We reduce the threshold used to compute the higher-education ratio and rely on the share of the population above 6 years old with at least primary education. Our results are not significantly affected by redefining the education ratio, with the exception of the education estimate, which is now statistically insignificant.

4.3. Interpretation of the results

These various robustness checks confirm the negative effect of immigration and the positive influence of market access and education on wages. Moreover, the structural derivation of the market access variable from theory provides us with a theoretical interpretation of its estimate: this figure corresponds to $\alpha_1 = 1/\sigma$, with σ being the elasticity of substitution between traded goods. As predicted, our estimates (Tables 1 and 2) of σ are greater than unity. Based on the GMM estimator, which controls for endogeneity of our explanatory variables, we obtain estimates of about 10. They are larger than recent estimates found in the United States context (Hanson, 2005). Based on a NEG model, Hanson's (2005) estimates of σ range between 4.9 and 7.6. However, higher elasticities of substitution are here expected since China trade more homogeneous and less differentiated goods than the U.S.

To compare the effects of market access and internal migration on wages, we test whether the difference between the two parameter estimates is statistically insignificant. Statistics (based either on χ^2 test for 2SLS or on F-test for GMM) are reported at the bottom of Tables 1 and 2. In all but one case (column 3 of Table 2), the tests suggest that in absolute value the coefficient on immigrant share is equal to the one on market access (at the 5% confidence level), in coherence with theory (see Eq. (13)).

Our results suggest that both migration and market access effects are statistically and economically significant. Another interesting result emerges, however. Based on the estimated year dummies, we find, holding other factors constant, that provincial wages increased on average by about 13.5% per year between 1995 and 2007.[20] This trend is common to all provinces and has been documented in the literature on shorter period of time (see earlier discussion and Adams et al., 2006; Lett and Banister, 2006). Four potential explanations can be advanced for this result. First, the annual growth in total factor productivity (TFP) may explain part of the national wage increase. Recent estimates suggest that China's TFP grew at an annual rate of 4% over the period 1993–2004 (Bosworth and Collins, 2008). Using a longer sample period (1984–2006), covering the more recent years that experienced a higher output growth, Li (2009) finds a slightly higher increase of TFP of 4.65%. Second, the national rise in service-sector prices may also explain part of the wage increase due to (partial) indexation.[21] Between 1995 and 2000, the national consumer price index of services has increased by 14.6% per year on average. After 2000, the price increase has been slower (e.g. between 2001 and 2007: 3.2% on average per year for residence services — including water, electricity and fuels; 3.3% for household services).[22] Third, social security reforms increasing minimum wages represent also part of the explanation. In 1993, the Chinese government began to reform the social security system and established a minimum wage in all China's provinces, with the exception of Tibet. In 2006, among all the provinces the highest minimum wages are found in three of our province-status super-cities: Shanghai (750 Yuan = US$ 94), Tianjin (660 Yuan = US 82.5) and Beijing (640 Yuan = US$ 80). The minimum wage is adjusted each year. Using data on these super-cities, we compute that minimum wages have increased on average by a factor 2.2 between 1995 and 2006. Finally, the increased integration of China into the World economy (e.g. China's opening up to trade and inward foreign direct investment) may explain part of the national wage increase. We control for the international market access evolution, but not for the recent increase in inward foreign direct investment (FDI). In 2005, China received about a third of the total inward FDI of all non-OECD countries. Using a panel data set of Chinese cities, Ge (2006) examines the impact of inward FDI on urban real wages. The results suggest that the existence of FDI has a significant and positive effect on urban wages and that this impact remains significant after controlling for other city characteristics.

Overall our estimates highlight that rapidly increasing wages in China correspond to a common national trend. Since total factor productivity growth appears to explain only one third of this trend, the China price has increased over the period. In the meantime, we find that migration flows have moderately slowed down wage increase in China. For the sake of illustration, we consider the coastal province of Guangdong which is one of China's most prosperous provinces. As of 2007, according to our data, Guangdong has the highest GDP among all provinces. It contributes approximately 12% of national GDP. Besides, Guangdong's average migrant share jumped from about 1.7% in 1995 to 5.3% in 2007. Based on our benchmark estimates (from column 5 of Table 1), a one-point increase in the immigrant share induces average wage to fall by approximately 2.4%. In Guangdong, this implies that immigration has slowed down wage growth by about 8.64% [$=(5.3-1.7)\times 2.4$] in total or 0.7% per year. In coastal provinces, where average immigration share has increased from 1.03 to 1.62%, this implies that immigration has slowed down wage growth by about 1.41% [$=(1.62-1.03)\times 2.4$] in total or 0.12% per year. The migration effect on wages has been, so far, moderate. However, with a further and larger relaxation of migration restrictions, as announced by the export-intensive provinces, inter-provincial migration may exert a stronger downward pressure on wages.

5. Conclusion

In this article, we examine the impact of access to markets and internal migration on Chinese provincial wages. We develop a wage equation from a New Economic Geography model to capture the upward pressure from national and international demand and downward pressure from migration. Using panel data at the province level, we find that migration has moderately slowed down Chinese wage increase over the period 1995–2007. With a further and larger relaxation of migration restrictions, we may expect a higher migration effect in the future and a greater erosion of Chinese wages growth.

Acknowledgments

We thank the editor Yves Zenou and two anonymous referees for many valuable comments which significantly improved the paper. We are grateful to Mary Amiti, Agnes Benassy-Quéré, Holger Breinlich, Andrew Clark, Thierry Mayer and Hylke Vandenbussche for detailed comments and fruitful discussions on this topic. We also thank Matthieu Crozet, Carl Gaigné, Kala Krishna, Miren Lafourcade, Daniel Mirza, Gianmarco Ottaviano, Farid Toubal and seminar participants at the University of Hong-Kong (ACE 2006), INRA Rennes, the University of

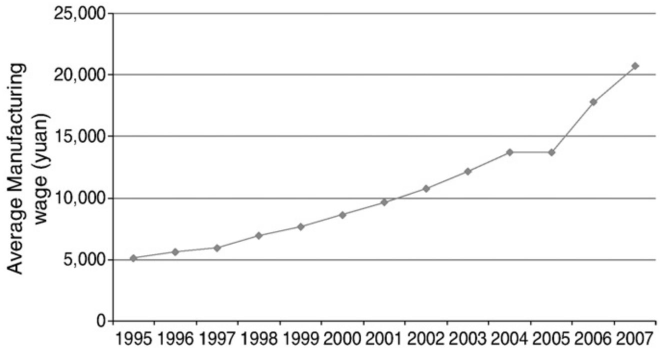
[20] To calculate this average, we first-difference the year dummy estimates and then compute the geometric mean of the antilog-transformed differences.

[21] Recall from Eq. (5) that we control for the manufacturing price index.

[22] After 2000, statistics on price of services have been decomposed in different categories. This renders the comparison before and after 2000 more difficult.

Ljubljana (EIIE 2007), the Central European University (EEA 2007), the University of Athens (ETSG 2007), the University of Louvain-la-Neuve, CERDI (University of Auvergne) and the University of Paris 1 for very helpful discussions and suggestions. This work has been supported by the ANR Jeunes Chercheurs.

Appendix A. Manufacturing wage growth (1995–2007)



Appendix B. Data description and sources

Our wage equation relies on various economic indicators constructed using China's Statistical Yearbooks National Bureau of Statistics (various years). They provide data at the provincial level, including province-status municipalities, on migration, population, climate, surface area, and average nominal wages for formal employees.

The immigrant share and the market access variables are defined in the core of the text (see Section 3).

The higher-education ratio variable is the share of population aged 6 and over that declares a secondary or above education attainment. This higher-education attainment is computed based on data collected in the yearly national sample survey on population change. It is the sum of people which attained junior secondary, senior secondary or college and higher.

Appendix C. Summary statistics, 1995–2007

	Obs	Mean	St. Dev.	Minimum	Maximum
Dependent variable					
Manufacturing wage	377	10,650	5561	3611	39,483
ln(manufacturing wage)	377	9.15	0.48	8.19	10.58
Explanatory variables					
Market access	377	0.026	.065	0.0002	0.441
ln(market access)	377	−5.24	1.67	−8.22	−0.82
Higher-education ratio	377	0.53	0.12	0.27	0.83
Rain	377	872.7	501.3	74.9	2678.9
Temperature	377	14.39	5.18	−7.8	25.4
Migration (tens of thousands)					
Newly arrived residents (1)	377	49.1	56.7	2.9	496.9
Newly arrived female residents (2)	377	23.4	26.6	1.7	243.6
Newly arrived male residents (3)	377	25.7	30.3	1.2	273.3
Population (4)	377	4333	2794	481	11,847
Immigrant share defined (in %) as					
(1)/(4)	377	1.21	0.76	0.26	5.32
(2)/(female population)	377	1.18	0.72	0.24	5.25
(3)/(male population)	377	1.24	0.82	0.22	5.73

Appendix C (*continued*)

	Obs	Mean	St. Dev.	Minimum	Maximum
Excluded instruments					
I_{temp}	377	15.49	2.33	10.24	21.17
I_{rain}	377	941.3	242.3	512.2	1959.7
15 year lagged population	377	3673.9	2457.6	373.7	10,998

Appendix D. Construction of market access

D1. Bilateral trade data

Different data sources are used to cover (i) intra-provincial (or intra-national), (ii) inter-provincial and (iii) international flows. These flows are all merged into one single data set to compute market access measures for each province based on their exports to all destinations (domestic and international).

D1.1. International data

- International trade flows in current USD come from IMF Direction of Trade Statistics.
- Intra-national trade flows of countries in current USD are computed following Wei (1996) as the difference between domestic industrial and agricultural production minus corresponding exports.
- Domestic industrial and agricultural production data for OECD countries come from the OECD STAN database. For other countries, the ratios of industrial and agricultural production as a percentage of GDP are extracted from Datastream. These are then multiplied by country GDP (in current USD) from World Development Indicators.

D1.2. Chinese data

- International trade flows of Chinese provinces are obtained from the Customs General Administration database, which records the value of all import and export transactions passing through customs. Provincial imports and exports are decomposed into up to 230 international partners (see Feenstra et al., 1998 and Lin, 2005).
- Inter-provincial trade is computed as trade flows between provinces and the rest of China and comes from input–output tables. Provincial input–output tables provide the decomposition of provincial production, and the international and domestic trade of tradable goods.[23]
- Intra-provincial flows of Chinese provinces are again computed following Wei (1996) as difference between domestic industrial and agricultural production minus corresponding exports.
- Domestic industrial and agricultural production data for Chinese provinces come from China's Statistical Yearbooks. Production in Yuan has been converted into current USD using the annual exchange rate. The exchange rate is the average exchange rate of the Yuan against the US dollar in the China Exchange Market and is obtained from China's Statistical Yearbooks.

[23] 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, i.e. between each province and the rest of China, in the total provincial trade is constant over time.

D2. Freeness of trade

The freeness of trade (ϕ_{ij}) is assumed to depend on bilateral distances ($dist_{ij}$) and a series of dummy variables indicating whether provincial or foreign borders are crossed.

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

We distinguish several different cases, according to whether i and j are provinces or foreign countries. This equation literally says that we allow for differentiated trade costs depending on whether bilateral trade occurs between a Chinese province and a foreign country ($-\delta \ln dist_{ij} - \varphi + \psi Contig_{ij}$), between two foreign countries ($-\delta \ln dist_{ij} - \varphi^{*} + \psi Contig_{ij}$), between a Chinese province and the rest of China ($-\delta \ln dist_{ij} + \vartheta$), within foreign countries ($-\delta \ln dist_{ij} + \xi$) and within Chinese provinces ($-\delta \ln dist_{ij}$). In the last two cases, only internal distance affects trade freeness. The accessibility of a Chinese province or a foreign country to itself is modeled as the average distance between producers and consumers in a stylized representation of regional geography, which yields $\phi_{ii} = dist_{ii}^{-\delta} = \left(2/3\sqrt{area_{ii}/\pi}\right)^{-\delta}$, where δ is the estimate of distance in the trade equation.

Note that being neighbors dampens the border effect ($Contig_{ij} = 1$ for pairs of partners which are contiguous) and that ζ_{ij} captures the unmeasured determinants of trade freeness, and is assumed to be an independent and zero-mean residual.

D3. Computation of market access

Table 3 reports the estimation results of the trade Eq. (17). Importer and exporter fixed effects are included in the regression so that the border effect for foreign countries ($-\delta \ln dist_{ij} + \xi$) is captured by their fixed effects. The reference category in the regression is within Chinese-province trade.

The estimates of distance and contiguity are in line with those in the related literature. We also confirm that the border effect inside China is important (Poncet, 2003). Furthermore, we find that impediments to trade are greater between China and the rest of the world than between the countries included in our sample (which are mostly members of the WTO and are therefore much more integrated into the world economy than was China in the 1990s).

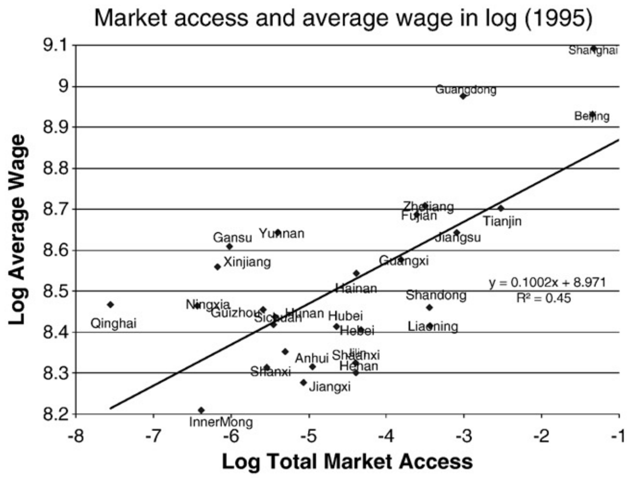
Trade equation estimates.

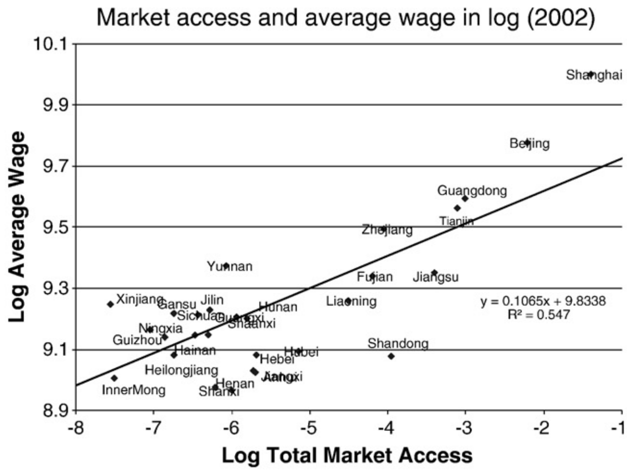
Columns	Dependent variable: ln(exports)				
	(1)	(2)	(3)	(2)	(3)
	1995	1999	2002	2005	2007
Exporter fixed effects	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes
Ln(Distance)	-1.15^{a}	-1.18^{a}	-1.24^{a}	-1.24^{a}	-1.23^{a}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Chinese	-5.85^{a}	-5.98^{a}	-5.37^{a}	-4.93^{a}	-4.57^{a}
Border Effect (B_{ij}^{f})	(0.26)	(0.25)	(0.25)	(0.27)	(0.28)
Foreign country	-3.07^{a}	-3.36^{a}	-2.37^{a}	-3.33^{a}	-3.22^{a}
Border Effect (B_{ij}^{f*})	(0.22)	(0.21)	(0.37)	(0.22)	(0.23)
Contiguity	1.92^{a}	1.85^{a}	1.90^{a}	2.01^{a}	2.09^{a}
	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)
Provincial	-2.82^{a}	-2.90^{a}	-2.37^{a}	-1.70^{a}	-1.39^{a}
Border Effect (B_{ij}^{c})	(0.36)	(0.38)	(0.37)	(0.43)	(0.45)
No. of Observations	23 386	24 750	26 249	27 240	27 554
R-squared	0.33	0.35	0.36	0.35	0.36

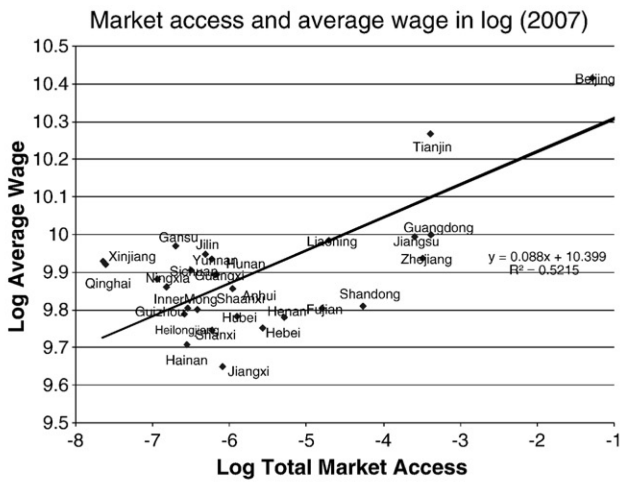
Heteroskedastic-consistent standard errors in parentheses, with [a] denoting significance at the 1% level.

D4. Market access and average manufacturing wage

The following figures plot the provincial market access as a function of average wage in log, separately for 1995, 2002 and 2007. We observe higher levels of market access for high-wage provinces. This is in line with the theoretical prediction of NEG models.







Appendix E. Econometric results

First-stage regressions.

Model of Table 1	Cols. (1) and (2)	Cols. (3) and (4)
	Dependent variable	
	Immigrant share	ln(population)
Excluded instruments		
I_{rain}	-0.00054^a	-0.00007^b
	(0.00017)	(0.00003)
I_{temp}	-0.225^b	-0.0097
	(0.098)	(0.0084)
15-year lagged population		-0.00007^b
		(0.00003)
Higher-education ratio	5.037^a	0.415^a
	(1.077)	(0.102)
ln(rain)	0.070	0.0066
	(0.085)	(0.0057)
Temperature	0.013^b	-0.0014
	(0.006)	(0.0012)
ln(population)	1.507	
	(1.416)	
ln(market access)	0.095	0.0136^b
	(0.909)	(0.0057)
R^2	0.557	0.737

Note: All regressions are first-stage of 2SLS province fixed effects regressions including year dummies. The sample size is 377. Clustered standard errors at the province level are reported in parentheses. [a], [b] and [c] denote significance at the 1%, 5% and 10% levels respectively. See text for more details.

Robustness test: long difference on manufacturing wage equation.

	Dependent variable: *ln*(manufacturing wage)		
Model	(1)	(2)	(3)
Included years	all (col. 1 Tab. 1)	1995, 1998, 2001, 2004 and 2007	1995, 2001 and 2007
Immigrant share	-0.132^b	-0.248^a	-0.249^c
	(0.067)	(0.096)	(0.150)
Market access	0.050^b	0.081^b	0.108^c
	(0.021)	(0.038)	(0.060)
Higher education	0.747^b	1.443^b	1.402
	(0.367)	(0.574)	(0.936)
Population	-0.582^b	-0.728^a	-0.735^a
	(0.237)	(0.265)	(0.262)
Rain	0.012	0.036	0.054
	(0.014)	(0.031)	(0.037)
Temperature	0.002	0.032	0.033
	(0.002)	(0.021)	(0.024)
Year 1996[i]	0.112^a		
Year 1997	0.144^a		
Year 1998	0.308^a	0.265^a	
Year 1999	0.379^a		
Year 2000	0.543^a		
Year 2001	0.642^a	0.610^a	0.639^a
Year 2002	0.745^a		
Year 2003	0.854^a		
Year 2004	0.991^a	0.929^a	
Year 2005	0.992^a		
Year 2006	1.241^a		
Year 2007	1.353^a	1.226^a	1.260^a
Observations	377	145	87
Within R^2	0.978	0.972	0.980
IV first-stage			
Migration R^2	2.7	4.9	4.0
Migration F-stat.	6.2^a	3.7^b	3.3^b
Weak test[j] χ^2	4.21	5.00	3.29
P-value	0.12	0.08	0.19
Overidentif. test[k] χ^2	0.32	1.53	1.79
P-value	0.57	0.21	0.18

Notes: Clustered standard errors at the province level are reported in parentheses. [a], [b] and [c] denote significance at the 1%, 5% and 10% levels respectively. See the text for details. [i]To save space, we do not report the constant and the clustered standard errors of the year dummies (they are available upon request). [j]Stock–Wright weak-instrument-robust inference test. [k]Hansen overidentification J statistic. Both test statistics are cluster-robust.

References

Adams, F.G., Gangnes, G., Shachmurove, Y., 2006. Why is China so Competitive? Measuring and explaining China's competitiveness. World Economy 29 (2), 95–122.

Amiti, M., Javorcik, B.S., 2008. Trade costs and location of foreign firms in China. Journal of Development Economics 85 (1–2), 129–149.

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics. Princeton University Press.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277–297.

Au, C.-C., Henderson, V., 2006. How migration restrictions limit agglomeration and productivity in China. Journal of Development Economics 80 (2), 350–388.

Baldwin, R., Forslid, R., Ottaviano, G., Martin, P., Robert-Nicoud, F., 2003. Economic Geography and Public Policy. Princeton University Press.

Bao, S., Bodvarsson, Ö.B., Hou, J.W., Zhao, Y., 2007. Interprovincial migration in China: the effects of investment and migrant networks. IZA Discussion Papers 2924.

Borjas, G.J., 2003. The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. Quarterly Journal of Economics 118 (4), 1335–1374.

Bosworth, B., Collins, S., 2008. Accounting for growth: comparing China and India. J. Econ. Perspect. 22 (1), 45–66.

Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: sorting matters! Journal of Urban Economics 63, 723–742.

Feenstra, R., Hai, W., Woo, W.T., Yao, S., 1998. The U.S.–China bilateral trade balance: its size and determinants. NBER Working Paper 6598.

Freeman, R., 2006. People flows in globalization. J. Econ. Perspect. 20 (2), 145–170.

Friedberg, R., 2001. The impact of mass migration on the Israeli labor market. Quarterly Journal of Economics 116 (4), 1373–1408.

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

Ge, Y., 2006. The effect of foreign direct investment on the urban wage in China: an empirical examination. Urban Studies 43 (9), 1439–1450.

Hanson, G., 2005. Market potential, increasing returns, and geographic concentration. Journal of International Economics 67 (1), 1–24.

Harris, C., 1954. The market as a factor in the localization of industry in the United States. Ann. Assoc. Am. Geogr. 64, 315–348.

Head, K., Mayer, T., 2006. Regional wage and employment responses to market potential in the EU. Regional Science and Urban Economics 36 (5), 573–594.

Hering, L., Poncet, S., 2010. Market access impact on individual wages: evidence from China. Review of Economic and Statistics 92 (1), 145–159.

Homer-Dixon, T.F., 1994. Environmental scarcities and violent conflict: evidence from cases. International Security 19 (1).

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

Lett, E., Banister, J., 2006. Labor costs of manufacturing employees in China: an update to 2003–04. Monthly Labor Review, November.

Li, K.-W., 2009. China's total factor productivity estimates by region, investment sources and ownership. Economic Systems 33 (3), 213–230.

Lin, S., 2005. Geographic location, trade and income inequality in China. In: Venables, A.J. (Ed.), Spatial Inequality and Development, Kanbur R. Oxford University Press, London.

Lin, J., Wang, G., Zhao, Y., 2004. Regional inequality and labor transfers in China. Economic Development and Cultural Change 52, 587–603.

National Bureau of Statistics of China, 2002. Figures on 2000 Population Census of China (CD-ROM). China Statistics Press, Beijing.

National Bureau of Statistics of China, China Statistical Yearbooks, People's Republic of China. Beijing, China Statistical Press, various years.

Naughton, B., 2007. The Chinese Economy: Transitions and Growth. MIT Press.

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

Poncet, S., 2006. Provincial migration dynamics in China: borders, costs and economic motivations. Regional Science and Urban Economics 36, 385–398.

Redding, S., Venables, A.J., 2004. Economic geography and international inequality. Journal of International Economics 62 (1), 53–82.

Roback, J., 1982. Wages, rents, and the quality of life. Journal of Political Economy 90 (6), 1257–1278.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Stock, J.H., Wright, J.H., 2000. GMM with weak identification. Econometrica 68 (5), 1055–1096.

Wang, F.-L., 2004. Reformed migration control and new targeted people: China's hukou system in the 2000s. China Quarterly 177, 115–132.

Wei, S.-J., 1996. Intra-national versus international trade: how stubborn are nations in global integration? NBER Working Paper 5531.

Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data Econometric. MIT Press, Cambridge.

WTO, 2008. International Trade Statistics 2008. Geneva.