Relaxing Hukou: Increased labor mobility and China's economic geography

Maarten Bosker a,b, Steven Brakman c,d, Harry Garretsen c,d,e, Marc Schramm f

Article history:
Received 17 February 2012
Available online 28 June 2012

JEL classification:
O18
R11
R12

Keywords:
China
Cities
Labour mobility
New economic geography

1. Introduction

China is one of the fastest growing economies in the world. Also, it is home to a huge (potential) internal market and a large number of very large and fast-growing cities. Arguably, China is a textbook case to analyze how agglomerating and spreading forces shape the economic landscape. To cite Krugman (2011, p. 6):

“surely the strong resemblance between China’s industrial clusters today and the industrial clusters of the 19th-century […] strikes a blow in favor of the argument for simple, common principles.”

The recent phenomenal growth record of China has been accompanied by migration from the western and central provinces to China’s eastern (coastal) provinces (World Bank, 2008), mainly involving rural migrants seeking jobs in China’s vibrant cities.1 As a result China’s urbanization rate has steadily risen over the last decades. Recent studies (e.g. Au and Henderson, 2006a,b), however, argue that Chinese cities are still undersized due to the severe restrictions on labor mobility that are imposed through the so called Hukou system (see also Chan, 2008, 2009). Compared to other countries in the world at a similar stage of economic development, China has a much more evenly sized city size distribution (Fujita et al., 2004, p. 2955), which arguably keeps it from reaping the full benefits of agglomeration.

In this paper, we study the impact of the Hukou system on the spatial distribution of economic activity in China. In particular, we provide answers to the question what a removal of the restrictions on labor mobility imposed by the Hukou system will imply for China’s internal economic geography. We do this by firmly basing ourselves in new economic geography (NEG) theory. In our view, NEG models are well suited to analyze the case of rapidly industrializing economies like China (see also Krugman, 2011). In particular, the focus in most NEG models on the agricultural and manufacturing sector may seem outdated for western countries but it is (still) relevant for a country like China. Also NEG stresses the relation between cities instead of treating cities like ‘floating islands in space’, as is common in urban economics (Fujita and Mori, 2005).

China’s Hukou system poses severe restrictions on labor mobility. This paper assesses the possible consequences of relaxing these restrictions for China’s internal economic geography. We base our analysis on a new economic geography (NEG) model. First, we estimate the important model parameters using data on 264 of China’s prefecture cities. Second, we use these estimates as inputs in a simulation of the full NEG model under different labor mobility regimes. We find that increased labor mobility leads to more pronounced core–periphery outcomes. Beijing, Shanghai, Guangzhou and Chongqing in particular will further strengthen their dominant place in China’s urban hierarchy. In addition, two other groups of cities can be distinguished: those in China’s populous economic heartland and more peripheral cities that are better shielded from competition with China’s economic heartland by virtue of their relative remoteness.

© 2012 Elsevier Inc. All rights reserved.

1 This paper is a substantially revised version of the CESifo Working Paper No. 3271 with the same title.
2 Corresponding author at: Faculty of Economics and Business, University of Groningen, The Netherlands.

E-mail address: j.h.garretsen@rug.nl (H. Garretsen).

1 Recently, there have also been reports that production is now actually moving away from coastal regions to more westward, non coastal regions (like the province Anhwei) where wage costs are lower (see “The Next China” in The Economist of July 31st 2010, pp. 46–48).

http://dx.doi.org/10.1016/j.jue.2012.06.002

0094-1190/ – see front matter © 2012 Elsevier Inc. All rights reserved.

0094-1190/ – see front matter © 2012 Elsevier Inc. All rights reserved.

0094-1190/ – see front matter © 2012 Elsevier Inc. All rights reserved.
In this paper, we focus on the NEG model introduced by Puga (1999). A nice feature of his model, that makes it particularly suitable for our case, is that the equilibrium spatial distribution of economic activity depends explicitly on the degree of interregional labor mobility. This makes the model a natural starting point to assess the relaxation of the Hukou restrictions on China's internal economic geography.

Our empirical strategy consists of two steps. First, based on a sample of 264 Chinese Prefecture cities, we estimate the equilibrium wage equation that is central in NEG models. This not only allows us to empirically establish the link between a city's wages and its market access as predicted by NEG theory (see also Hering and Poncet, 2010a); more importantly it provides us with the key structural model parameters that are crucial in our second step.

It is this second step that makes our study different from recent related NEG studies for China (e.g. Ma, 2006; Lin, 2003; Hering and Poncet, 2010a,b; De Sousa and Poncet, 2011; or Amiti and Javorcik, 2008). Compared to these studies, where the estimation of the NEG wage equation is the central objective (hereby implicitly taking the spatial allocation of labor (and firms) as given), we go one step further and make use of the complete NEG model. Using our estimates of the important model parameters as input, we simulate the full NEG model under various different regimes of labor mobility. This allows us to make model-based predictions of what the consequences of a relaxation of China's Hukou system could mean for its internal economic geography.

We find that increased interregional labor mobility will lead to more pronounced core–periphery outcomes. In the resulting much more uneven city-size distribution, China's internal demand will be a key determinant of its future economic geography. Beijing, Shanghai, Guangzhou and Chongqing in particular will further strengthen their dominant place in China's urban hierarchy. In addition, two other groups of cities can be distinguished: those in the populous heartland of China offering preferential access to China's enormous internal market, and several more peripheral cities that are shielded from competition with China's largest population centres by virtue of their substantial distance from these centres increasing both migration costs for workers and trade costs for firms.

2. Labor mobility in China and the implications from a NEG perspective

2.1. The Hukou system

Since the 1950s the Chinese authorities have been much concerned with internal labor migration flows and rural–urban labor migration in particular. The Chinese government alternated between periods of more and less restrictive migration policies (Zhao, 2005, Fujita et al., 2004; Chan and Buckingham, forthcoming; World Bank, 2008, chapter 5, and Chan, 2009), but ever since the 1950s the so called Hukou system has been a major feature of its internal migration policies. Although the system has become less restrictive over time it is still very much a prominent characteristic of the Chinese labor market (see i.e. the detailed description of the Hukou system in The Economist of May 6th, 2010). The Hukou system is equivalent to an internal visa arrangement that is meant to regulate migration. In recent decades the system has been quite restrictive, not only by limiting migration flows from rural to urban areas, but also by putting a brake on inter-urban migration flows (Chan and Buckingham, forthcoming; Au and Henderson, 2006a,b; Poncet, 2006; Henderson, 2009; Zhang and Zhao, 2011).

Without a visa for a particular location, a Chinese citizen has no or only limited rights to housing, sell property, education, food or social security in that location. Those rights are tied to one's official place of residence and a change in residency (if a citizen for instance would try to move from a rural area to a city) will only be matched with a transfer of these rights if the (local) authorities hand out a visa or permit for the new place of residence.

Until recently (see in particular Chan, 2009 or Chan and Buckingham, forthcoming, pp. 13–14), migration under the Hukou system had two equally important dimensions. The first concerns the restrictions on access to education, health care, housing, etc. allowed to above. To be entitled with the aforementioned rights to public provisions, households need a local Hukou. The second dimension is functional and refers to the distinction between agricultural and non-agricultural workers. Non-agricultural workers with a local Hukou are traditionally entitled to more rights than agricultural workers. Combined, the local and functional classification define four possible categories of residents (Chan, 2009, p. 202).

In recent years, the Hukou system has become less of a constraint (Chan and Buckingham, forthcoming). With urban wages outstripping rural wages, the result has been an increase in (temporary) migration into the booming cities. One important change is that currently, in many cases, the local instead of the central government decides upon the permits. This gives local governments some degree of freedom with respect to the leniency of granting these permits. A second change is that the distinction between a non-agricultural and an agricultural Hukou has become relatively less important (Chan, 2009). The key issue for a migrant is now foremost whether he or she has a local Hukou, and thus whether he or she has access to public provisions in the destination location.

Although the system has become less stringent and more differentiated, the Hukou system is still very restrictive. “all these restrictions sharply reduce the benefits and raise the costs of migration, particularly into large cities. Migration is limited and most migration is short-term, or “return” migration. (…) Overall the Hukou system holds hundreds of millions of people in locations where they are not exploiting their earnings potential.” (Fujita et al., 2004, p. 2957; see also, Zhang and Zhao, 2011). Chan (2008) estimates that between 1982 and 2006, the annual volume of Hukou migrants remained very stable at about 17–20 million people (see also Fujita et al., 2004). This stability suggests that neither the scrapping of the functional Hukou, nor the decentralization of the migration policy has had a substantial impact on the flow of permanent migrants to the cities.

2.2. The effects of the Hukou system from an economic geography perspective

What are the consequences of the Hukou system for China? First, China's degree of urbanization is smaller than expected based on its level of economic development. Fig. 1 illustrates this line of argument in a stylized way (Henderson, 2009):

This lower than expected urbanization rate, at least partly arises...
embraced urbanization, they are still wary of mega-cities.\textsuperscript{7} The Hukou system restricts people to take full advantage of their productive capabilities by holding them back in areas where they cannot fully exploit these capabilities (see also Zhang and Zhao, 2011). Moreover, others have argued that the agglomeration rents associated with urbanization are underutilized (see also Fujita et al., 2004).

Second, Au and Henderson (2006a,b) forcefully argue that Chinese cities are too small as a consequence of the migration restrictions.\textsuperscript{7} The Chinese city size distribution is characterized by many fast-growing small-to-medium sized cities, its largest cities have not grown as fast.

Apart from the general prediction that relaxing Hukou’s migration restrictions will most likely result in higher urbanization rates and larger and/or more uneven-sized Chinese cities (Au and Henderson, 2006a), little work has been done on assessing the impact of increased labor mobility on the internal economic geography of China. Will it result in very strong core–periphery patterns, with most economic activity concentrated in a few very large cities? And if so, which cities or regions will be able to attract labor mobility quite easily results in full agglomeration (see also Bosker et al., 2010 or Whalley and Zhang, 2007). The Hukou system restricts people to take full advantage of their productive capabilities by holding them back in areas where they cannot fully exploit these capabilities (see also Zhang and Zhao, 2011). Moreover, others have argued that the agglomeration rents associated with urbanization are underutilized (see also Fujita et al., 2004).

Before we discuss the basics of the model, we note that there is one crucial difference between our model and the original Puga (1999) model. This concerns the inclusion of a non-tradable services sector in our model, i.e. here called housing (see Helpman, 1998, and Hanson, 2005). The inclusion of a housing sector changes the mix of agglomeration and spreading forces. By increasing the cost of living as a city grows, it introduces an additional spreading force (see also Puga, 1999, p. 324, footnote 19), decreasing the likelihood of ending up in the unrealistic scenario of complete agglomeration.\textsuperscript{8}

Fig. 1. Urbanization and gross national income (GNI) per capita for countries.

\textsuperscript{7} Or as The Economist of September 18th 2010 states (p. 84): “China’s small cities exploded in number. But its biggest metropolises conspicuously failed to explode in size (...). This partly reflects a conscious policy. Although China’s rulers have embraced urbanization, they are still wary of mega-cities.”

\textsuperscript{8} Without this additional spreading force, the model with perfect interregional labor mobility quite easily results in full agglomeration (see also Bosker et al., 2010 or Head and Mayer, 2004, pp. 2652–2653).
In case of intersector labor mobility within Chinese Prefecture cities is important. Fujita et al. (2004, p. 2958, Table 2) e.g. document that in the 1990s the non-agricultural employment growth outstripped population growth (which can be interpreted as a sign of inter-sector labor mobility).

The agricultural good is produced using land and labor (combined in a Cobb–Douglas production function, with \( 0 < \sigma < 1 \) denoting labor’s share in agricultural production).\(^6\) Its market is assumed to be perfectly competitive with free entry and exit. Moreover it is freely tradable between regions. The assumed production structure in agriculture implies decreasing returns to labor in this sector, so that any attempt of manufacturing firms to lure workers away from the agricultural sector implies wage increases.\(^1\) In this sense the fixed stock of land \( K_i \) acts as a second spreading force in the model. Together with the fixed housing stock, \( H_i \), it captures in a stylized way the costs of congestion associated with larger agglomerations.

The industrial sector produces heterogeneous varieties of a single good under monopolistic competition and free entry and exit, incurring so-called ‘iceberg’ trade costs when shipped between regions (\( t_y \geq 1 \) goods have to be shipped from region i to let one good arrive in region j). Industrial production technology is characterized by increasing returns to scale. The production input is a Cobb–Douglas composite of labor and intermediates, with \( 0 \leq \mu \leq 1 \) the Cobb–Douglas share of intermediates. Intermediates enter the production function as a composite manufacturing good that is specified as a CES-aggregate (with \( \sigma > 1 \) the elasticity of substitution across varieties) of all manufacturing varieties produced. Finally, regions are allowed to differ in their production efficiency, \( c_i \), yet firms within the same region are assumed to be similar.\(^12\)

Consumers have Cobb–Douglas preferences over: the agricultural good (\( A \)); the non-tradable service, housing, (\( H \)); and a CES-composite (also with \( \sigma > 1 \) the elasticity of substitution across varieties) of manufacturing varieties (\( M \)), with \( 0 \leq \gamma_j \leq 1 \), \( s = M, A, H \) the non-tradable share of each (aggregate) good (with \( \gamma_A + \gamma_H + \gamma_M = 1 \)).

Next, equilibrium (factor) prices and demand follow from profit and utility maximization on behalf of firms and consumers respectively, perfect intersector labor mobility, and free entry and exit of firms (Appendix A for the details). Subsequently, firms and people (only the former when labor is interregional immobile) move across regions in search for higher profits and real wages until no one has an incentive to change his/her location. Puga (1999) shows that the corresponding equilibrium distribution of firms and people critically depends on the assumptions regarding interregional labor mobility.

In case of no labor mobility between regions, the model reduces to the following four equilibrium conditions, that determine (1) the (manufacturing) price index \( q_i \); (2) nominal wages \( w_i \); (3) total expenditures on manufactures \( e_i \); and (4) the price of the non-tradable service (housing) \( p_{hi} \) in each region i:

\[
q_i = \left( \frac{1}{1-\mu} \sum_j (c_j L_j q_j^{-\mu} c_j^{-1} p_{hi}^{-1+\sigma})^\frac{1}{1-\sigma} \right)^\frac{1}{\sigma} \tag{1}
\]

\[
w_i = q_i^\mu \left( \sum_j (c_j L_j q_j^{-\mu} c_j^{-1} p_{hi}^{-1+\sigma})^\frac{1}{1-\sigma} \right)^{-\frac{\mu}{1-\sigma}} \tag{2}
\]

\[
e_i = \gamma_A Y_i \tag{3}
\]

\[
h_i p_{hi} = \gamma_H Y_i \tag{4}
\]

where in addition to the variables and parameters already introduced, \( c_j \) denotes the share of region i’s labor force in manufacturing and \( r(w_i) \) the rent earned per unit of land in region i.

Eq. (1) concerns the CES price index for manufacturing varieties, which in general becomes smaller the closer varieties are produced in the neighborhood (low transport costs) of i. Eq. (2) denotes the nominal wage equation and it is derived from the market equilibrium conditions (supply equals demand; because of mark-up pricing the equilibrium price equation can be re-formulated as a wage equation). In general, firms can afford to pay higher wages if they are close (low transportation costs) to markets, and intermediate firms are located close to final production. Eq. (3) defines expenditure on manufactures, which depends on wage income and landowner rents (again see Appendix A for a detailed description). Finally, Eq. (4) is the equilibrium condition on the housing market. Given that the housing stock is assumed to be fixed, the price of houses increases the higher \( \gamma_H Y_i \).

With interregional labor mobility, nominal wage equality between the A and M sector within each region is no longer sufficient to ensure equilibrium. Workers now also move in response to real wage differences between regions. As a result, with perfect interregional labor mobility, equilibrium is characterized not only by (1)–(4) but also by real wage equalization across all regions:

\[
\omega_i = q_i^{-\sigma} p_{hi}^{1-\sigma} w_i = \omega_0, \quad \forall i \tag{5}
\]

Note that our addition of a non-tradable service sector, housing, has no consequences when labor does not move between regions. Because of the simple way in which the housing market is modeled (therefore it is a fixed housing stock), the equilibrium condition on the housing market (4) simply implies that house prices are higher the more income \( \gamma_H Y_i \) spent on housing. Without interregional labor mobility, Eq. (4) helps to determine the housing price, but the housing price cannot influence worker’s location decisions (labor is immobile between regions).

This changes when labor moves freely between regions. Real wages (5) not only depend on the price index of manufactures \( q_i \), but also housing prices \( p_{hi} \). When more workers move to region i, \( Y_i \) will increase and, given the fixed stock of housing and the fixed expenditure share on housing, this implies an increase in housing prices. The latter will, ceteris paribus, decrease real wages in region i, decreasing its attractiveness to future migrants (or even making people leave the region in search of higher real wages elsewhere). Therefore the fixed supply of housing acts as a spreading force that increases as a city grows.

Consequently, the model gives very different predictions regarding the equilibrium spatial distribution of people and firms depending on the degree of interregional labor mobility. With labor completely immobile between regions, we will see firms moving to places offering better profit prospects, but people are, by definition, not moving around. When on the other hand, people are able to move around, the equilibrium distribution of people will change (except in the (very) unlikely case that real wages

\(^6\) Puga (1999) defines the agricultural sector somewhat more general. However, in order to derive analytical results, a Cobb-Douglas production function in agriculture is used, see Puga (1999, p. 318).

\(^7\) See also Chan (2000, p. 208) on the wage elasticity of intra-regional manufacturing labor supply.

\(^8\) These differences arise from e.g. differences in human capital, resource endowments, etc, determining why some regions may be able to offer higher wages.
are initially already equalized), with people moving towards places offering higher real wages. Whether or not this will result in more or less agglomeration, is a priori not clear. This depends on the important model parameters such as trade costs, the relative importance of housing, and manufacturing and agriculture in people’s utility: but also on regions’ initial endowments of arable land, population, and housing stock.

The Chinese Hukou system poses considerable restrictions on the degree of interregional labor immobility. Theory predicts that things will almost certainly change substantially when these restrictions are alleviated. However, what exactly will happen is a priori unclear. This is exactly what we will go after in the next two sections. In Section 5, we obtain estimates of all the important model parameters. Next, in Section 6, we use the estimated model parameters, in combination with China’s current distribution of people, housing stock and arable land, as inputs to simulate (using the full NEG model) what China’s economic geography might look like under different labor mobility regimes.

4. Data set and a first look at China’s economic geography

Before turning to our main empirical exercise, we provide a brief overview of China’s current economic geography. At the highest level of aggregation, China is composed of 33 administrative units (22 provinces, 5 autonomous regions, 4 large municipalities (Beijing, Shanghai, Tianjin, and Chongqing) and 2 special regions (Hong Kong and Macau). The 2nd tier of regional division, the so-called Prefecture level, in China consists of 333 regions. Of these 333 regions at the Prefecture level, 283 regions are Prefecture-cities.13 Our data set covers a large subset of these Prefecture cities for the period 1999–2005: 264 of the 283 cities are included. Even though the data do not cover the whole of China, it captures most of its population and economic activity. Our 264 prefecture cities cover 86% of total population in China and 96% of total GDP. In Appendix B we list for every administrative unit14 the corresponding Prefecture cities for which we have data. Appendix B also provides a map of China based on the above mentioned administrative division of China.

The data on these 264 Chinese prefecture cities come from the Chinese Data Center at the University of Michigan (see http://chinadatcenter.org/newcdc/). The original data source is the National Bureau of Statistics of China (NBS). From this database we have data on each prefecture city’s expenditures (income), total population (both urban as well as rural), employment, arable land (in km²), as well as distance to the nearest major ports, secondary education enrolment, the share of employment in banking and finance, and its total area. Moreover we calculated (great-circle) distances between each prefecture city-pair.

Our data in principle covers the years 1999–2005. There is, however, some doubt in the literature about the reliability of the data in non-Census years that is reported by local and regional governments, particularly for the population data. In Section 5, we therefore also present estimation results when only using the 2000 Census data. Moreover, we use the 2000 Census data as input for all our simulations in Section 6.

Figs. 2a and b illustrate China’s economic geography for the year 2000. Fig. 2a shows that the largest concentration of population can be found in the Eastern part of China.15 This area includes some of China’s largest cities (e.g. Beijing, Shanghai, and Guangzhou) but also non-coastal provinces like Honan, Huphe, Anhwei, Kiangsu, and Siantung. Taken together these five provinces alone are home to 29% of the Chinese population and to 34% of the total population in our 264 Prefecture cities. This presence of quite a few, relatively populous, Prefecture cities in China’s (non-coastal) heartland will turn out to be important in the long-run equilibrium analysis of relaxing the Hukou system in Section 6.

Fig. 2b additionally shows that the distribution of manufacturing activity does not correspond one-to-one with the distribution of people. Secondary industry employment (panel b) is more spatially concentrated (Herfindahl index (HI) = 0.011) than population (HI = 0.004).16 Beijing, Shanghai or Chongqing, the three largest cities in terms of population (together accounting for 5% of total population) are home to an even larger 12% of secondary industry employment. The five earlier mentioned provinces in China’s populous heartland together contain about 27% of total secondary industry employment.

5. Obtaining estimates of the important NEG model parameters

To assess how China’s economic geography, as depicted by Fig. 2, might change due to a loosening of the Hukou restrictions, we first need to get estimates of the important parameters of our NEG model (see Section 3). Estimates of the share of intermediate inputs (μ) in manufacturing production, and the share of income spent on manufactures (γM) are obtained from a regional input–output table for China for 2000. We take μ = 0.511 (1–O table: Chinese intermediate demand for Chinese manufacturing by Chinese manufacturing firms), and γM = 0.343 (1–O table: Chinese final demand for Chinese manufacturing as share of total final demand for Chinese output).17 From the Statistical Yearbook of the Chinese Bureau of Statistics we obtain the Cobb–Douglas share of labor in agriculture production as the share of wages in total agricultural value added, θ = 0.879 (the high value indicates that agricultural production is (still) relatively labor intensive in China). Finally, we base the share of income spent on housing on Ye and Wang (2008) and set it at 0.118.18

5.1. Estimating the wage equation

We also need an estimate of σ, the elasticity of substitution between manufacturing varieties. This parameter is not so easily obtained from the raw data. To get it, we structurally estimate a log-linearized version of the equilibrium wage Eq. (2):

\[\ln(w_{ij}) = \frac{1}{\sigma} \ln \left( \sum_{q=0}^{\infty} \epsilon_{ij}^{\sigma} (1 - \epsilon_{ij}^{\sigma})^{-1} \right) + \alpha_1 + \alpha_2 + \alpha_3 \ln X_{ij} + \epsilon_{ij}\]

where the productivity differences between regions, ci, in (2), are captured by several observed control variables (Xci) (see below for more on our specific choice of controls) and a random error term εij that is assumed to be uncorrelated with the included regressors.

In addition, we allow ci to differ between cities by including

13 Note that the prefectures not only include the urban population of a city, but also the rural area surrounding a particular city, in many Prefecture cities the data set the majority of the population as well as the bulk of the land area is classified as non-urban. Apart from the 283 Prefecture cities the 2nd tier of regional administration also consists of 17 Prefectures (mainly in Xinjjang and Tibet), 30 Autonomous Prefectures (in western China, regions with a large share of ethnic minorities) and 3 Leagues (regions in Inner Mongolia, see http://en.wikipedia.org/wiki/Administrative_divisions_of_China#Prefecture_level_subdivisions).

14 We have no data on the following 3 administrative units: Hong Kong, Macau, and Tibet. The same underlying data set is also used by for instance Au and Henderson (2006a) or Moreno Monroy (2011).

15 As we will explain in Section 6 the Prefecture “city” of Chongqing (the large black area in the middle of China) is a rather special case. See footnote 27 in particular.

16 A perfectly even spread of population or secondary employment would give a HI of 0.0038.


18 As pointed out by one of the referees, this share may not accurately reflect the opportunity cost of housing. In robustness checks we will also use a higher housing share of 0.25 which is more comparable to the international standard.
prefecture city dummies (captured by the $x_i's$ in (6)), and we include year dummies to control for general improvements in production efficiency affecting all cities in the sample equally (captured by the $x_i's$ in (6)).

Estimating the NEG wage equation lies at the heart of most earlier empirical NEG studies. It has been used to verify one of NEG’s main predictions that market access (the term between brackets) plays an important role in explaining interregional differences in nominal wages. In case of China, Hering and Poncet (2010a) for example estimate a similar wage equation as (6). They provide strong evidence in support of NEG’s prediction that workers living in regions with better market access earn higher wages. A few issues need special attention when one estimates (6). First, note that (6) is actually a simplified version of (2). Besides the market access term, (6) should also contain supplier access, $q^{\nu^\mu-1}$. As shown and emphasized by Redding and Venables (2004), including both supplier and market access in the wage equation easily leads to multi-collinearity problems. Following Redding and Venables (2004) or Hering and Poncet (2010a), we exclude this term when estimating the wage equation (effectively assuming that $\mu = 0$).

Second, in order to limit problems with reverse causality (see Hanson, 2005 for more details on this), we include a market access measure in (6) that is based on data at one level of aggregation higher than our prefecture data. These provincial variables ($j = 1, \ldots, 30$) are constructed by aggregating the corresponding prefecture variables in the sample to the province level (Hanson, 2005) uses a similar aggregation for his sample of US counties.

Third, in the absence of actual trade costs data, we, as all other existing empirical NEG studies, proxy trade costs by specifying a trade cost function (see Bosker and Garretsen, 2010 for a recent discussion on the use of a trade cost function in empirical NEG). In particular, we proxy trade costs by a simple power distance function: $T_{ij} = (D_{ij})^\delta$, where $D_{ij}$ is defined as the distance between city $i$ and the capital of province $j$. Trade costs within one’s own province ($T_{ii}$) are proxied similarly using $a$, by now standard, measure of internal distance: $D_{ii} = \frac{2}{\sqrt{area}}$. The use of such a distance function introduces the distance decay parameter $\delta$ as an additional important model parameter for our simulations in the next section.

Finally, we do not directly observe the manufacturing price index, $q$, that is part of the market access term. To overcome this problem we follow Brakman et al. (2004) and approximate $q$ as follows. Wages are one of the most important determinant of prices. This enables us to proxy a province’s manufacturing price index by a weighted combination of a province’s own wage level and the average wage level outside that province (weighted by distance to correct for the transport costs involved in importing goods from other provinces): $q_j = \frac{1}{3} \left( \frac{1 - \delta}{\rho_0} / C_0 + \frac{1}{\rho_0} / C_0 \right) D_{ij}$

where $w$ is the average wage of all other provinces in the sample, $D_{ij}$ is the distance between the capital of province $j$ and the nearest economic centre; Beijing, Shanghai or Guangzhou, and $\delta$ the share of employees in province $j$. When estimating (6) we substitute (7) for $q_j$ in Eq. (6). We prefer this approximation over the alternative method used by Hanson (2005) that would require the assumption that real wages are already equalized across Chinese cities.

In the literature, two basic empirical strategies have been used to estimate (6). The first, introduced by Redding and Venables (2004), is to use the information contained in bilateral trade data to construct a theory-based measure of each region’s market access. Subsequently this constructed measure of market access is included in the estimation of the wage equation. In case of China, Hering and Poncet (2010a,b), Lin (2003) and Ma (2006) have followed this strategy using data on inter-provincial trade. The second strategy, following Hanson (2005), is to estimate (6) directly using nonlinear estimation techniques. We opt for this second strategy here because there is, to our knowledge, no bilateral trade data available at the Prefecture-city level that (sufficiently) covers our sample. Other related market access studies for China, adopting a

---

19 Another option to deal with unavailability of data on city price indices is to go for nominal instead of real market access (see Au and Henderson 2006a), yet this takes us out of the world of NEG.

20 This alternative involves information on Housing prices. If we approximate housing prices, $p_{hi}$, for each city $i$, by the price of floor space per m² and substitute (5) into the wage equation (6), where the housing prices $p_{hi}$ become part of the market access term, the manufacturing price index $q$ drops out. Using the otherwise similar specification as in column (a) in Table 1, we can then estimate the substitution elasticity, $\sigma$ and the distance parameter $\delta$. Interestingly, this alternative estimation also provides one with an (indirect) estimate of the share of income spent on housing $\tau_{hi} = 0.12$ which is close to the 0.118 value we now use in our simulations based on Ye and Wang (2008).
similar direct approach as we do are e.g. Amiti and Javorcik (2008), and Au and Henderson (2006a).

Table 1 shows the results of estimating (6). In all our regressions, we use prefecture city specific GDP per capita as our measure of wages $w_i$. Data on the average wage of workers in the urban part of each prefecture city, and data on the total wage of employees are available, but following Hering and Poncet (2010b) who argue that the available prefectural city wage data do not sufficiently reflect wages in the private sector, we did not use either of these two wage measures as our main dependent variable, opting for gdp per capita instead. Expenditures $e_i$ are proxied by a prefecture city’s income.

Finally, to capture possible productivity differences between prefecture cities that also determine the observed wage differences between cities (i.e. the $c_i$ in (2)) we include several additional control variables ($X_i$) in (6). First, we add population density. A city’s density is related to higher productivity levels because of increasing returns associated with the well-known Marshallian externalities of e.g. labor market pooling and knowledge or input sharing. Second, we include a proxy for human capital as a higher educated workforce is likely to be more productive (see also Hering and Poncet, 2010b; Breinlich, 2006 and Bosker and Garretsen, forthcoming). Third, to control for the possible relevance of the economic structure of a city, we add the share of banking and finance in total Prefecture-city employment to our regression.

In addition to these controls, we also include time, and prefecture city dummies respectively to control for unobserved (fixed) determinants of a city’s wage. Prefecture fixed effects in particular capture the idea that some cities have a comparative advantage in attracting workers and/or industry. In particular it controls for the fact that some cities in the sample are located along China’s coast and are often specialized in exporting to world markets (Hu, 2002; Au and Henderson, 2006a), therefore paying higher wages than other cities. City fixed effects to a large extent controls for this, supply driven, part of China’s economic geography. Also, the fixed effects capture the effect(s) of Chinese industrial policy where specific regions or cities are favoured over others.

A final note before discussing our results concerns the fact that market access, depicted as in (6), only captures access to China’s internal market. China is, however, not a closed economy and market access with respect to the rest of the world may also be relevant. It is straightforward to incorporate this in the market access term: the summation in (6) is extended to also include the rest of the world (see Hering and Poncet, 2010a). In doing so, we take the rest of the world as consisting of the following three economic blocs: the USA, Japan and the EU.22

Table 1 shows the results of our estimations. Column (a) shows our main panel estimation (using yearly data over the 1995–2005 period) including a full set of city- and year-specific fixed effects. Our findings confirm those by earlier NEG based wage studies for China: we find strong evidence that market access plays an important role in explaining the observed wage differences between Chinese prefecture cities. Most relevant for our purposes is that both the substitution elasticity $\sigma = 5.886$ and the transport cost parameter $\delta = 0.632$ are significant. The estimated value for $\sigma$ implies a market access (MA) coefficient [1/$\sigma$, see (5)] of about 0.17: that is, a 1% increase in a city’s market access is associated with a wage increase of 0.17%. This is in line with Hering and Poncet (2010a, MA-coef. $\approx 0.1$), Hering and Poncet (2010b, MA-coef. $\approx 0.07$) and Moreno Monroy (2011, MA-coef. $\approx 0.25$). The transport cost parameter likewise is in line with earlier studies, see e.g. Au and Henderson (2006a) who, following Poncet (2006), put the distance coefficient at 0.87.

The results on the control variables largely confirm earlier findings by e.g. Hering and Poncet (2010a), Breinlich (2006) or Bosker and Garretsen (2012). Human capital is an important determinant of wages. Furthermore, population density contributes significantly (and positively) to wages, suggesting that the benefits of agglomeration still outweigh any negative congestion effects for the average Chinese city. The share of banking and finance employment is not significant at the 5% level.

As an important robustness check to our findings, we did a separate cross-section estimation of the wage Eq. (6) using data for the Census year 2000 only.23 As already mentioned in Section 4, doubts have been raised on the reliability of non-Census year data. Column (b) shows that focussing on the 2000 Census data only, results in

---

22 As a robustness check on the results shown in column (a), we also included the economic “mass” outside of China as a separate regressor to (6). To take account of differences between prefecture cities in their ease of access to the coast we also interacted this measure with a city’s distance to the nearest seaport. The results are essentially the same as those shown in column (a). Since the specification underlying column (a) is closest to the underlying NEG model we prefer these estimation results. Note that distance to the nearest seaport itself is captured by the prefecture fixed effects. Similarly the distance weighted sum of foreign GDP is captured by the year fixed effects.

23 For their sample of 51 Chinese cities, Hering and Poncet (2010a) use micro data which allows for a much better control of various wage determinants besides market access.
very similar estimates for the substitution elasticity and the trade cost parameters compared to those obtained using our full-sample. Interestingly, we now also find the expected negative effect of being located further away from one of China’s seaports (this effect is absorbed by the prefecture city fixed effects in column (a)25).

As discussed in Section 3, the NEG model we use distinguishes between two different cases of labor mobility: complete interregional labor immobility or perfect interregional labor mobility.29 In the former case, workers do not migrate: an (extreme) Hukou-scenario. The latter case is the extreme case of completely abandoning the Hukou system: workers migrate in response to the smallest wage differential between cities [see Eq. (5)]. These simple migration dynamics between locations can be represented by:20

\[ \frac{d\lambda_i}{dt} = \psi(\lambda_i - \bar{\lambda}) \]  

where \( \bar{\lambda} = \sum_j \lambda_j \) is the average real wage across all locations and \( \lambda_i \) is the share of people living in region \( i \); \( \psi \) is a parameter determining the speed of adjustment. The more real wages in region \( i \) differ from average real wages, the faster people leave or move to (depending on real wages being higher or lower than average real wages respectively) region \( i \).

Both cases are too stylized in our view, although defensible from a theoretical point of view in order to keep things simple. The assumption that not a single person moves between cities does not capture the current Hukou state of the world. Also, the assumption that people migrate in response to the slightest wage differentials does not realistically reflect a world without Hukou’s restrictions. We will refer, in the simulations, to these two extreme scenarios, and include more realistic cases. First, and most easily done, we simply take China’s economic geography in 2000 as our benchmark Hukou scenario. This situation is shown in Fig. 2 above. As discussed earlier, this situation is based on the most reliably available Census data. It reflects the state of China’s economic geography under the current Hukou restrictions on labor mobility. Next, starting from this situation in 2000, we simulate what happens if the Hukou restrictions were to be relaxed. In doing so, we introduce more realistic migration dynamics than simply assuming that people are willing to move (anywhere) in response

see Section 3, the housing market acts as a spreading force in case of interregional labor mobility, so that we need starting values for the city-specific housing stock \( H_i \). Since data on housing stock are lacking, we approximate \( H_i \) by taking the city’s (2000) urban population share in total Chinese urban population as a proxy of its share in the overall Chinese housing stock.

### 6.2. Allowing for different interregional labor mobility regimes

As discussed in Section 3, the NEG model we use distinguishes between two different cases of labor mobility: complete interregional labor immobility or perfect interregional labor mobility.29 In the former case, workers do not migrate: an (extreme) Hukou-scenario. The latter case is the extreme case of completely abandoning the Hukou system: workers migrate in response to the smallest wage differential between cities [see Eq. (5)]. These simple migration dynamics between locations can be represented by:20

\[ \frac{d\lambda_i}{dt} = \psi(\lambda_i - \bar{\lambda}) \]  

where \( \bar{\lambda} = \sum_j \lambda_j \) is the average real wage across all locations and \( \lambda_i \) is the share of people living in region \( i \); \( \psi \) is a parameter determining the speed of adjustment. The more real wages in region \( i \) differ from average real wages, the faster people leave or move to (depending on real wages being higher or lower than average real wages respectively) region \( i \).

Both cases are too stylized in our view, although defensible from a theoretical point of view in order to keep things simple. The assumption that not a single person moves between cities does not capture the current Hukou state of the world. Also, the assumption that people migrate in response to the slightest wage differentials does not realistically reflect a world without Hukou’s restrictions. We will refer, in the simulations, to these two extreme scenarios, and include more realistic cases. First, and most easily done, we simply take China’s economic geography in 2000 as our benchmark Hukou scenario. This situation is shown in Fig. 2 above. As discussed earlier, this situation is based on the most reliably available Census data. It reflects the state of China’s economic geography under the current Hukou restrictions on labor mobility. Next, starting from this situation in 2000, we simulate what happens if the Hukou restrictions were to be relaxed. In doing so, we introduce more realistic migration dynamics than simply assuming that people are willing to move (anywhere) in response

see Section 3, the housing market acts as a spreading force in case of interregional labor mobility, so that we need starting values for the city-specific housing stock \( H_i \). Since data on housing stock are lacking, we approximate \( H_i \) by taking the city’s (2000) urban population share in total Chinese urban population as a proxy of its share in the overall Chinese housing stock.
to the slightest wage differential. In particular, our migration dynamics differ from those depicted by (5) and (8) in two important ways.

First, we know from the migration literature in general and from a few China specific studies in particular (see e.g. Zhang and Zhao, 2011), that workers do not migrate as soon as they can earn only a Yuan or two more in another location. Instead, migration is costly (not only in monetary terms, but also in social terms of having to leave behind family, friends, etc.). People only migrate when the wage differential between two locations is large enough to outweigh these costs. To take this into account, we assume that a spatial equilibrium is not characterized by Eq. (5), but by:

\[
\max[(\omega_i - \omega_j)/\omega_j] = x^\%\quad \forall i,j
\]

where \(x\) is the minimum wage differential at which people are no longer willing to move to other cities offering higher real wages. Based on findings by Zhang and Zhao (2011) we assume a minimum wage differential threshold of 15% in our baseline simulation runs (i.e. \(x^\% = 15\)%).

Second, a recent study by Zhang and Zhao (2011) on rural–urban migration in China finds that migrants ceteris paribus prefer short distance migration. There exists a significant income-distance trade-off. These findings corroborate Poncet (2006) who shows that Chinese migration patterns depend negatively on distance, and have a strong provincial bias (see also Fujita et al., 2004). Given these earlier studies, we take the empirical findings by Poncet (2006) seriously, and adapt the stylized migration dynamics in (8) as follows:

\[
d_{ij} = \sum_j m_{ij} \quad \text{with} \quad m_{ij} = \begin{cases} 
\omega_i^{0.87} \omega_j^{0.13} & \text{if} \omega_i > \omega_j \\
\omega_i^{0.87} \omega_j^{0.13} & \text{if} \omega_i < \omega_j
\end{cases}
\]

where \(m_{ij}\) denotes the change in a city’s share in total population due to immigration from another lower wage city \(j\), or due to emigration to another higher wage city \(j\) (depending on whether real wages in city \(j\) are lower or higher respectively). \(NB_{ij}\) is a dummy variable indicating whether cities \(i\) and \(j\) are located in the same province.

6.3. China’s internal economic geography when labor mobility is increased

Having discussed the main inputs and migration dynamics into our model simulations, we are now in a position to make NEG-based predictions regarding China’s internal economic geography under different interregional labor mobility regimes.

Before we present the results of our model simulations, we would like to stress that these model simulations are certainly not aimed at giving a pin-point prediction about what will happen to Chinese agglomeration patterns once more labor mobility is introduced. The NEG model that we use (or any theoretical model for that matter), is too stylized to capture all the essentials of China’s economic geography (see our discussion in Sections 1–3), and to expect spot-on predictions about which cities will “gain or lose” and by how much. Still, we think that the model-based approach is able to give insightful predictions about the type of places in China that will more likely see a population in- or outflow in case the Hukou restrictions were alleviated.

Although the details differ (and will be discussed below), the upshot of the model simulations with interregional labor immobility is that they all indicate that a relaxation of the Hukou system would lead to much stronger core–periphery patterns. Even though suffering from the general NEG tendency to overstate the degree of agglomeration, the simulations point towards the following qualitative findings.

Fig. 3 shows the resulting spatial equilibrium in our baseline “no Hukou”-scenario (i.e. \(x^\% = 15\)% in (9)). Compared to the baseline 2000 Hukou case shown in Fig. 2a, population and industry are both much more concentrated (\(H_{\text{final}} = 0.082\) compared to 0.011 in 2000, and \(H_{\text{popul}} = 0.047\) compared to 0.004 in 2000). Only 52 Prefecture cities still have a positive share of firm and worker population. Within this group, three categories of cities can be distinguished. First, there are the four largest cities: Beijing, Shanghai, Guangzhou and Chongqing. These four cities are already today among China’s largest. Loosening the restrictions on labor mobility will only reinforce their dominant position: in equilibrium these four cities house a third of total population. Market and supplier access are high in these four, initially already very large agglomerations. As a result, they can offer relatively high wages and profits that attract workers and firms alike. In this respect it is very interesting to note that these four cities are also today’s four top migration destinations in China identified by Zhang and Zhao (2011).

A second group of cities is located in the area demarcated by the “Big 4”; roughly speaking the ‘diamond’ area indicated by the dotted lines in Fig. 3. This area comprises some of China’s most populated provinces, such as Sajantung, Fujian, Kwantang, Anhwei, Honan, or Hunan. These 31 cities (together they comprise 41% of population in equilibrium) are all part of the populous heartland of China. They benefit from increased labor mobility because of their prime location in the centre of China’s economic geography. Compared to the “Big 4”, it is not their (initial) size that is decisive, but the easy access to nearby markets and cities. They do face competition for people and firms from the attractive agglomeration economies offered by the “Big 4”, yet ‘survive’ in equilibrium because of their lower cost of housing, which at some point stops people from being willing to incur the migration costs needed to move to e.g. Shanghai or Beijing. Also, compared to the more peripheral cities, they can more cheaply import goods produced in one of the four main cities.

The third group of cities are initially (in terms of 2000 population) large but peripheral cities (i.e. outside the ‘diamond’ shown in Fig. 3) such as Harbin in the Heilunkiang province in northern China, or Kunming in the western province Yunnan. In equilibrium these 17 cities contain around 25% of total population. Despite being peripheral, these cities ‘survive’ the relaxation of the Hukou restrictions and are able to resist the attraction of the more central cities. On the one hand the economic centres in South – South Eastern China are simply too far away for people to consider migrating to them (given the costs incurred if they were to do so). On the other hand footloose firms also do not leave the periphery because these firms are shielded from competition from the economic

34 NEG models for example tend to over- or under-predict the degree of agglomeration – see e.g. Bosker et al. (2010).
35 Hereby also supporting the notion that the large(st) Chinese cities are still
36 Again it should be stressed that the spatial equilibrium visualized in Fig. 3 is not meant as an actual prediction of the spatial allocation of economic activity when interregional labor mobility is allowed for. To do that our model remains too stylized; however these results in our view do serve as a reliable indication of what the qualitative changes in China’s economic geography when abandoning the Hukou system (i.e. will this lead to more agglomeration or not, and if so where is this most likely).
centre of China by transport costs. An extreme example of such a peripheral, stand-alone, city is Urumqi in the far north-western corner of China.\footnote{When focussing on “Han-China” only (i.e. we exclude the Inner Mongolian Autonomous Region and the Xinjiang Uyghur Autonomous region, as well as the western-most prefecture in Gansu province), we find essentially the same results. The correlation between the equilibrium population distribution with and without “Han-China” is 0.994. In equilibrium we are left with three fewer peripheral cities: Urumqi, Baotou and Chifeng; and with one more central city: Jingshui in Hupeh province.}

How sensitive are these results for the wage threshold \(x\) in Eq. (9)? Put differently, how does the spatial equilibrium depend on the level of migration costs that we assume in our simulations? Irrespective of the wage threshold used, the qualitative pattern as discussed above remains unchanged in the sense that the same three groups of cities can be distinguished, with the four largest cities always very dominant (Beijing, Shanghai, Guangzhou and Chongqing). As expected, the higher the threshold, the more cities survive. A threshold of 10%, 15%, 20%, results in 23, 52, and 120 cities, respectively. Not surprisingly, the 23 cities that continue to exist with a 10% threshold are a subset of the 52 cities remaining with a 15% threshold, which is itself a subset of the 120 cities remaining with a 20% threshold. Also, the less costly migration, the more dominant the “Big 4” become.\footnote{In the clearly unrealistic case of workers moving anywhere over any distance in response to the smallest wage differential (i.e. migration dynamics as defined by (5) and (8)), we end up in an, also clearly unrealistic, equilibrium with only Beijing surviving as China’s sole “giga”-city.} They house half the population with a 10% threshold, and 16% of population with a 20% threshold (compared to 33% in our baseline). Interestingly, the relative importance of central and peripheral cities existing in equilibrium also changes. With a 20% threshold 75 central and 41 peripheral cities remain, comprising 57% and 27% of total population respectively. In the 10% threshold scenario instead, a similar number of peripheral and central cities remain (9 vs. 10), housing a respective 27% and 22% of total population. As such our model predicts that the more peripheral cities (e.g. Harbin or Kunming) will be able to better withstand the ‘lure of the Big 4’: the centrally located cities in particular find it increasingly difficult to hold their populations as migration costs decrease.

As an additional robustness check, we also verified how our addition of the housing market to the Puga (1999) model affects our results. As discussed in Section 3 this addition adds an additional spreading force to the model that, ceteris paribus, should decrease the likelihood of ending up in a spatial equilibrium characterized by full agglomeration. Indeed, when simulating our model without the housing sector (but still including the more realistic migration dynamics as in (9) and (10)), we always end up in a more agglomerated spatial equilibrium than if a housing sector were included (i.e. \(HI = 0.092\) compared to \(HI = 0.047\) in our baseline scenario, also only 21 cities remain in equilibrium compared to 52 in our baseline scenario). Without the housing sector, the cost of living does no longer increase in growing cities. In our main simulation results housing costs do increase faster in relatively faster growing cities, putting a stronger cap on agglomeration pressures resulting in the more evenly spread spatial equilibrium shown in Fig. 3.

If we increase the share of income spent on housing to 0.25 (which is more in line with the international standard than the 0.118 that we use in our baseline simulations), we end up in a less agglomerated spatial equilibrium, i.e. \(HI = 0.050\). In this case, 78 cities remain compared to the 52 in our baseline scenario. What is interesting is that the Big 4’s share in overall population is hardly affected; it drops 1 ppt from 33.4% to 32.4%. The biggest effect of increasing the spreading force posed by the housing market is that it changes the share of population remaining in the central and peripheral cities. Compared to the baseline scenario the central cities’ share in total population drops 3.3 ppt and that of the peripheral cities rises by 4.4 ppt. The main driver of this shift is that many more peripheral cities remain when increasing the spreading force posed by the housing market.

Overall, taking the full NEG model seriously leads to interesting insights about the (possible) future of China’s internal economic geography, were its government to relax the current Hukou restrictions on people’s mobility. Our simulation results show that

Fig. 3. China’s economic geography: relaxing Hukou.
China's internal geography is likely to be substantially affected by such a policy change. Corroborating claims by e.g., Au and Henderson (2006a), our simulations predict that relaxing Hukou's restrictions will result in much more pronounced core–periphery patterns. This may not be that surprising, but our simulations also show that the initial differences in population or market size do not necessarily need to be conclusive in determining which cities are most likely to grow and which ones to shrink once interregional labor mobility is allowed for.

We find that increased labor mobility will substantially strengthen the dominant position of today's largest cities in China (Beijing, Shanghai, Guangzhou and Chongqing). But, an important qualitative insight from our results is that these "Big 4" other cities remain as well. These cities are not necessarily among China's largest today. One group of cities ‘survives’ based on its central location with respect to China's currently most populated provinces. Another, more peripheral group of cities is also able to retain population and economic activity. These cities retain their population by virtue of their relative peripheral location which increases migration costs for workers on the one hand and trade costs for firms on the other hand, effectively shielding them from competition with China's main centres of production.

7. Conclusions

The Chinese Hukou system puts severe restrictions on people's mobility. It has been argued to be one of the reasons why China is relatively underurbanized considering its level of economic development, as well as why China's city-size distribution is much more evenly spread than in other countries. In this paper we provide theory-based insights into the question how China's internal economic geography will be affected by a relaxation of the Hukou restrictions on people's ability to move between cities.

To do this, we firmly base our analysis a new economic geography (NEG) model that, in our view, is a very good starting point to analyse the relationship between market access, interregional labor mobility and agglomeration. Our paper differs from related NEG studies for China (e.g., Ma, 2006; Lin, 2003; Hering and Poncet, 2010a,b; De Sousa and Poncet, 2011; Amiti and Javorcik, 2008). In contrast to these papers, we do not take the spatial allocation of labor (and firms) as given, by focussing exclusively on the equilibrium (nominal) wage equation. Instead, we make use of the complete NEG model. Based on estimates of the key structural NEG model parameters, we simulate the consequences of relaxing China's Hukou system, making use of the full set of equilibrium conditions implied by the model. By simulating the spatial equilibrium of the full NEG model under different scenarios of interregional labor mobility, we are able to make predictions about what a relaxation of the Hukou system may imply for the future distribution of people and economic activity within China.

Our main findings show that a relaxation of the restrictions on labor mobility posed by the Hukou system will lead to more pronounced core–periphery outcomes. Our analysis shows that China's own internal market will be a particularly important determinant of its future economic geography in case the Hukou system was to become less restrictive. In the resulting much more uneven city-size distribution, Beijing, Shanghai, Guangzhou and Chongqing in particular will further strengthen their dominant position in the Chinese urban hierarchy. However, besides these "Big 4", other cities remain as well. Two groups of cities can be distinguished: those in the populous heartland of China offering preferential access to China's internal market, and several more peripheral cities that survive because their distance from China's main population centres shields them from competition with China's largest population centres.

Acknowledgments

We thank William Strange and two anonymous referees for very useful comments and suggestions. Also we thank seminar participants at the London School of Economics, University of Cambridge; the 2009 ETGS Conference in Rome; Utrecht University; University of Groningen, the 2009 NARSC conference in San Francisco, the EBRD in London; and especially Bart van Ark, Vivian Chen, Bernard Fingleton, Vernon Henderson, Laura Hering, Ron Martin, Ana Moreno Monroy, Henry Overman, Pierre Picard, Sandra Poncet, Mark Roberts, and Tuo Shi for their comments and suggestions. Special thanks go to Ana Moreno Monroy and Zhao for their help in composing the figures.

Appendix A

A.1. The multi-region version of the Puga (1999) model including a housing sector

Consider a world consisting of R regions. Each of region \(i = 1, \ldots, R\) is populated by \(l_i\) workers, endowed with a stock of non-tradable services (e.g., Housing) \(H_i\) and endowed with \(K_i\) units of arable land. Each region's economy consists of two sectors: agriculture and industry. Labor is used by both sectors and is mobile between sectors within a region and it is either mobile or immobile between regions. Land on the other hand is used only by the agricultural sector and is immobile between regions. The fixed housing stock is owned by absentee landlords. As we also explained in Section 3 of the paper, the difference between our model and the multi-region model as developed by Puga (1999) is the inclusion of a housing sector.

A.2. Production

The agricultural good is produced under perfect competition and free entry and exit using Cobb-Douglas technology. Moreover it is freely tradable between regions. The industrial sector produces heterogeneous varieties of a single good under monopolistic competition and free entry and exit. Industrial production technology is characterized by increasing returns to scale, i.e., production of a quantity \(x(h)\) of any variety \(h\) requires fixed costs \(c_x\) and variable costs \(c_h\) that are both assumed to be the same in each region, but can differ between regions due to differences in e.g., production efficiency, \(c\).

This production structure, together with free entry and exit and profit maximization, ensures that in equilibrium each variety is produced by a single firm in a single region. The production input is a Cobb-Douglas composite of labor and intermediates, with \(0 < \mu < 1\) the Cobb-Douglas share of intermediates. Intermediates enter the production function as a composite manufacturing good that is specified as a CES-aggregate (with \(\sigma > 1\) the elasticity of substitution across varieties) of all manufacturing varieties produced.

Firms in principle sell their goods to all regions. But, shipping their goods to foreign markets incurs so-called ' iceberg' trade costs \((\tau_p > 1)\) goods have to be shipped from region \(i\) to let one good arrive in region \(j\). Taking these costs into account, gives the following profit function that is similar for each firm in region \(i\):

\[
\pi_i = \sum_{q=1}^{Q} p_q(h) x_q(h) / \tau_p - w_m^{\sigma} q^\mu c + \beta \sum_{R}^{K} x_k(h) \]  \hspace{1cm} (A.1)

where \(p_q(h)/\tau_p\) is the price of a variety produced in country \(i\), \(q\) is the price index of the composite manufacturing good, \(w_m^{\sigma}\) the manufacturing wage in region \(i\).

The inclusion of \(c\) rationalizes that we control for efficiency differences between prefecture cities in our estimation of the wage equation.
A.3. Preferences

Consumers have Cobb–Douglas preferences over the agricultural good \(A\); the non-tradable service, housing, \(H\); and a CES-composite (also with \(\sigma > 1\) the elasticity of substitution across varieties) of manufacturing varieties \(M_i\), with \(0 \leq \gamma_i \leq 1\). In this model, it increases with a region’s land endowment and with the Cobb–Douglas share of each (aggregate) good (with \(\gamma_M^s + \gamma_A + \gamma_H = 1\)). Specifying the composite manufacturing good this way ensures demand from each region for each manufacturing variety, which, together with the fact that each variety is produced by a single firm in a single region, implies that trade takes place between regions.

A.4. Equilibrium

Having specified preferences as well as the production technologies of the manufacturing and agricultural good, the equilibrium conditions of the model can be calculated. Profit maximization and free entry and exit determine the share of labor employed, \(L^M_i\), and the wage level, \(w^M_i\), in agriculture, as well as the rent earned per unit of land \(r^M_i\). The former two in turn pin down the share of workers in manufacturing, \(\zeta_i\). Given the assumed Cobb–Douglas production function in agriculture, with labor share \(\theta\), we have that:

\[
\zeta_i = \frac{L^M_i}{L_i} = 1 - \frac{L^A_i}{L_i} = 1 - K_i \left( \frac{\theta}{w^A_i} \right)^{\frac{1}{\beta}}
\]

where \(0 < \theta < 1\) denotes the Cobb–Douglas share of labor in agriculture, and \(L^M_i\) and \(L^A_i\) the number of workers in manufacturing and agriculture respectively. Eq. (A.2) shows that, in contrast to Krugman (1991), where agriculture uses only land \((\theta = 0)\), or to Venables (1996), where agriculture employs only labor \((\theta = 1)\), the share of a region’s labor employed in manufacturing is endogenously determined in this model. It increases with a region’s labor endowment and agricultural wage level and decreases with a region’s land endowment and with the Cobb–Douglas share of labor in agricultural production. Consumer preferences in turn determine total demand for agricultural products in region \(i\) as:

\[
X^A_i = (1 - \gamma_M^s - \gamma_H)Y_i
\]

In the industrial sector, profit maximization with free entry and exit gives the familiar result that all firms in region \(i\) set the same price for their produced manufacturing variety as being a constant markup over marginal costs:

\[
p_i = \frac{\sigma\beta}{\sigma - 1} c_i q_i^M w_i^{M^{1-\sigma}}
\]

where \(q_i\) is the price index of the composite manufacturing good in region \(i\) defined by:

\[
q_i = \left( \frac{1 - \mu}{\sigma - 1} n_i p_j^{1-\mu} \right)^{\frac{1}{\mu}}
\]

where \(n_i\) denotes the number of firms in region \(i\) and

\[
W^M_i = \left[ (1 - \mu)n_i p_i \left( \frac{\sigma - 1}{\sigma\beta} (\alpha + \beta x_i) \right) \right] (\zeta_i L_i)^{-1}
\]

\(W^M_i\) is the manufacturing wage level in region \(i\).

Utility maximization on behalf of the consumers in turn gives total demand for each manufacturing variety produced (coming from both the home region \(i\) as well as foreign regions \(j\)) which is the same for each variety in the same region due to the way consumer preferences are specified:

\[
x_i = \int p_i^a e_q^j q_j^{1-\mu} c_i^\alpha t_i^\sigma
\]

where in (A.7) demand from each foreign region \(j\) is multiplied by \(\tau_j\) because \((\tau_j - 1)\) of the amount of the product ordered from region \(i\) melts away in transit (the iceberg assumption).

\[
e_i = \gamma_A Y_i + \mu n_i p_i \left( \frac{\sigma - 1}{\alpha\beta} (\alpha + \beta x_i) \right)
\]

is total expenditure on manufacturing varieties in region \(i\) (the first term representing consumer expenditure and the second term producer expenditure on intermediates), where

\[
Y_i = \gamma_A Y_i + w_i^M (1 - \zeta_i) L_i + w_i^H \zeta_i L_i + r^M_i K_i + n_i \pi_i = Y_i
\]

is total consumer income consisting of spending on housing, workers’ wage income, landowners’ rents and firms’ profits respectively (note again, that housing rents are assumed to be earned by absentee landlords so they do not enter this income equation). Due to free entry and exit these profits are driven to zero, which (after substituting for wages, thereby uniquely defining a firm’s equilibrium output at:

\[
x_i = \alpha(\sigma - 1)/\beta
\]

Finally, to close the model, the labor markets are assumed to clear:

\[
L_i = L^M_i + L^A_i
\]

where the demand for labor in agriculture, \(L^A_i\), follows from the assumption of Cobb–Douglas technology in agriculture and the term between square brackets represents the total manufacturing wage bill. Moreover, equating labor supply to labor demand in the industrial sector gives an immediate relationship between the number of firms and the number of workers in industry:

\[
n_i = \frac{\zeta_i L_i}{\alpha(\sigma - 1) q_i^M w_i^{M^{1-\sigma}}}
\]

A.5. Long run equilibrium

To solve for the long run equilibrium, Puga (1999) distinguishes between the case where labor is both interregionally and intersectorally mobile and the case when it is only intersectorally mobile. Without interregional labor mobility, long run equilibrium is reached when the distribution of labor between the agricultural and the industrial sector in each region is such that wages are equal in both sectors. This is ensured by labor being perfectly mobile between sectors driving intersectoral wage differences to zero. When instead labor is also interregional mobile, not only intersectoral wage differences are driven to zero in all regions in equilibrium. Workers now also respond to real wage (utility) differences between regions by moving to regions with the higher real wages (utility) until real wages are equalized between all regions, hereby defining the long run equilibrium.
A.6. Interregional labor immobility

The long run equilibrium in case of interregional labor immobility can now be shown to be a solution \( \{w_i, q_i\} \) of the equilibrium equations that have to hold in each region. In our case (when using wage-worker space) these are, using the fact that in equilibrium \( w^M_i = w^S_i = w_i \),

\[
q_i = \left( \frac{1}{1-\mu} \sum_j (\gamma L_j q_j^{-\sigma} c_j^{\sigma} w_j^{1-\sigma} q_j^{1-\sigma}) \right)^{1/(1-\sigma)} \tag{A.13}
\]

\[
w_i = q_i^{\mu/(\mu-1)} c_i^{1/(\mu-1)} \left( \sum_j q_j^{1-\sigma} (1-\sigma) \right)^{1/(\mu-1)} \tag{A.14}
\]

\[
e_i = \gamma M_i \frac{Y_i}{\gamma H_i} (w_i L_i + K_i f(w_i)) + \mu/(1-\mu) w_i c_i L_i \tag{A.15}
\]

where (A.13) is obtained by substituting (A.4) and (A.12) into (A.5), (A.14) by substituting (A.4) and (A.10) into (A.7) and (A.15) by substituting (A.4), (A.9), (A.10), and (A.12) into (A.8).\(^40\) The final equilibrium condition concerns the housing market:

\[
H_{t+1} = \gamma H_i Y_i \tag{A.16}
\]

With \( p_{t+1} \) is the price of the non-tradable service (housing) in region \( i \), \( H_i \) is the fixed housing stock and \( \gamma H_i Y_i \) is the share of income spent on housing in region \( i \). In this case, that is without interregional labor mobility, Eq. (A.16) helps to determine the housing price but housing price cannot influence worker’s location decisions. It is with interregional labor mobility, see below, that housing prices start to matter for the long run equilibrium.

A.7. Interregional labor mobility

With interregional labor mobility, workers will move between regions in response to real wage differences until the interregional real wage differences, which are possible to persist when workers are unable to move between regions, are no longer present. More formally, the long run equilibrium solution \( \{w_i, q_i\} \) for each region \( i \) has to adhere to the additional condition that real wages, \( \omega_i \), are equal across all regions:

\[
\omega_i = q_i^{-\gamma M i} p_H^{-\gamma H} w_i = \omega_i, \quad \forall i \tag{A.17}
\]

So with interregional labor mobility we have equilibrium conditions (A.13)-(A.17). Compared with Puga (1999) or the core new economic geography model by Krugman (1991), our model with interregional labor mobility shows two main differences. The first one is the inclusion of the housing market Eq. (A.16). The second one relates to the fact that the real wage equalization condition (A.17) not only contains the price index of manufactures \( q_i \) but also the housing prices \( p_H \). When more workers move to region \( i \), \( Y_i \) will increase and, given the fixed stock of housing and the fixed expenditure share on housing, this will imply an increase in housing prices. The latter will ceteris paribus decrease real wages in region \( i \), thereby acting as a spreading force that gets stronger as more workers move to this region.

Appendix C

\(^40\) Also, without loss of generality, \( \alpha \) and \( \beta \) are set at \( 1/\sigma \) and \( (\sigma - 1)/\sigma \) respectively.
Appendix B. Chinese provinces and the prefecture cities in our data set

<table>
<thead>
<tr>
<th>Kwantung</th>
<th>Sjantung</th>
<th>Honan</th>
<th>Szechwan</th>
<th>Ahnwei</th>
<th>Liaoning</th>
<th>Kiangsu</th>
<th>Hunan</th>
<th>Hupeh</th>
<th>Kwangsi</th>
<th>Heilungkiang</th>
<th>Hebei</th>
<th>Tsckiang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guangzhou (4)</td>
<td>Shangouan</td>
<td>Qingdao (C)</td>
<td>Zibo</td>
<td>Luoyang (C)</td>
<td>Panzhihua</td>
<td>Shenyang (C)</td>
<td>Fuzhou (P)</td>
<td>Wuhan (C)</td>
<td>Changsha (C)</td>
<td>Nanning</td>
<td>Harbin</td>
<td>Shijiazhuang (C)</td>
</tr>
<tr>
<td>Zhubai</td>
<td>Shantou</td>
<td>Dongying</td>
<td>Anyi</td>
<td>Shantou</td>
<td>Huaihai</td>
<td>Tianjin</td>
<td>Hebei</td>
<td>Hebei</td>
<td>Hebei</td>
<td>Hebei</td>
<td>Hebei</td>
<td>Hebei</td>
</tr>
<tr>
<td>Zhanjiang (C)</td>
<td>Hezhou</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
<td>Xiamen</td>
</tr>
<tr>
<td>Meizhou</td>
<td>Shanwei</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
<td>Heyuan</td>
</tr>
<tr>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
<td>Yangjiang</td>
</tr>
<tr>
<td>Qingxiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
<td>Zhanjiang</td>
</tr>
<tr>
<td>Dongguan</td>
<td>Heze (C)</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
<td>Zhumadian</td>
</tr>
<tr>
<td>Zhongshan</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
<td>Chaozhou</td>
</tr>
<tr>
<td>Jeyang</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
<td>Hainan</td>
</tr>
<tr>
<td>Yunfu</td>
<td>Haikou</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
<td>Sanya</td>
</tr>
<tr>
<td>Tsinghai</td>
<td>Beijing</td>
<td>Tianjin</td>
<td>Shanghai</td>
<td>Chongqing</td>
<td>Chongqing</td>
<td>Chongqing</td>
<td>Chongqing</td>
<td>Chongqing</td>
<td>Chongqing</td>
<td>Chongqing</td>
<td>Chongqing</td>
<td>Chongqing</td>
</tr>
</tbody>
</table>

Notes: Bold faced names denote provinces. (4), (C) and (P) relate to Figs. 3 (plus subsequent discussion); they denote “Big 4”, Central and Peripheral cities respectively.
References


