The Geography of Talent in China During 2000-2015: An Eigenvector Spatial Filtering Negative Binomial Panel Data Approach

Abstract: The increase in China's skilled labor force has drawn much attention from policymakers, national and international firms and media. Understanding how educated talent locates and re-locates across the country can guide future policy discussions of equality, firm localization and service allocation. Prior studies have tended to adopt a static cross-national approach providing valuable insights into the relative importance of economic and amenity differentials driving the distribution of talent in China. Yet, how the influence of these factors has evolved over time remains unexplored. Recently released official statistical data now enables space-time analysis of the geographic distribution of talent and its determinants in China. Using four-year city-level panel data from national population censuses and 1% population sample surveys conducted every five years between 2000 and 2015, we examine the spatial patterns of talent across Chinese cities and their underpinning drivers evolve over time. Our results revealed that the spatial distribution of talent in China is persistently unequal and spatially concentrated between 2000 and 2015, although the extent of this unevenness and concentration has decreased slightly over time. Results also showed that economic opportunities have remained main attraction and retention forces of talent in main cities, but that key local amenities have also played a role in shaping the residential decisions of talent, particularly urban public services and greening rate. These results highlight that China's talent's settlement patterns over the first fifteen years of the 21st century have been mainly driven by urban economic opportunities *versus* amenity-related factors.

Keywords: talent; spatial pattern; determinants; eigenvector spatial filtering; panel data analysis; China

1. Introduction

Human capital accumulation has been recognized as a vital driving force in the development of knowledge-based economies (Lucas, 1988; Romer, 1990). As a result, policymakers have been universally concerned with ways of attracting and retaining human talent (Florida, 2002). In China, the national number and share of the highly educated population as a share of the labor force have both rapidly increased over the last fifteen years. The 2015 national 1% population sample survey reported that 170.93 million Chinese citizens held a college degree or above, an increase of 72.4% from the same sample collected in 2000. The rise of skilled individuals in China's labor force has drawn much attention from policymakers, employers, and the media. China's government has placed increasing emphasis on enlarging the national pool of highly skilled laborers by developing *The National Medium and Long-Term Talent Development Plan (2010-2020)* in 2010 (State Council of China, 2010). This plan states that China needs to transition from labor-intensive to talent-intensive economic activities to increase its competitiveness in the global economy. Similarly, the report of the 19th National Congress of China in 2017 pointed out that China should firmly implement "*the strategy of reinvigorating China through human resource development*". It also states that China should cultivate a large number of strategic talent, leading talent in the fields of science and technology, young talent, and highly innovative teams with international standards. Over the last 10 years, local governments at all levels have been engaged in the competition for skilled workers on the global stage, implementing policies to attract and retain talent, such as the "Peacock Plan" (*kong que ji hua*) in Shenzhen (Shenzhen Municipal Government, 2019). These policies include housing allowances, fast-track hukou transfers, and lower thresholds for bank loan applications.

There has been a surge in the number of studies on the migration and redistribution of various kinds of talent in China over the past decade (Florida et al., 2012; Gu et al., 2019a; Liu & Shen, 2014a, 2014b; Liu et al., 2017; Liu & Xu, 2017; Nie & Liu, 2018; Qian, 2010; Zeng et al., 2019). Due to data availability, prior studies have tended to adopt a static cross-national approach providing novel insights into the relative importance of economic and amenity differentials driving the migration patterns of talent. Yet, changes in the influence of these factors on shaping shifts in the spatial dynamics of migration over time remain underexplored. Understanding these changes is key to guide evidence-based economic development policies targeted at attracting and retaining talent. Additionally, the geographical scale of existing studies has been relatively coarse (i.e. provincial level), restricting our understanding of the location of talent across the urban-rural continuum within and between provinces. Moreover, prior empirical studies tend to focused on understanding the migration and redistribution patterns of highly educated talent during the period pre-2010. Yet less is known about the contemporary patterns of Chinese highly educated talent post-2010. The availability of the 2010 census and 2015 1% sample survey data enables investigating more recent spatial patterns of talent in China and identifying their key drivers.

More generally, while previous analysis on migration in Europe has identified a pattern of highly significant spatial autocorrelation in the distribution of talent (Miguélez et al., 2010; Rodríguez-Pose & Tselios, 2011), little has been done to model these spatial effects in a regression framework explicitly. It is shown that talent stock in a particular region may be positively related to the pool of talent in its surrounding areas, which leads to spatial autocorrelation. Researchers have discussed the sources for the spatial autocorrelation: they observed that talent gathered in nearby regions for lower communication cost, knowledge spillover, and sharing public resources (Fujita & Thisse, 1996; Miguélez et al., 2010). Interregional frequent mobility of talent between neighboring areas may also lead to positive spatial autocorrelation in talent stocks (Liu & Shen, 2017), which may also result from close cross-regional social talent networks (Gu et al., 2019a). A handful of studies have used autocorrelation analysis to investigate the distribution of talent within a country or a region (Nie & Liu, 2017; Sone et al., 2016). Yet little research has taken into account the presence of spatial dependence when assessing the geographic distribution of talent within countries. Without proper treatment, latent spatial autocorrelation may lead to estimation bias in regression models. Eigenvector spatial filtering (ESF) is a technique for capturing latent spatial autocorrelation in model residuals (Liu & Shen, 2017; Gu et al., 2019b). Each eigenvector extracted from a spatial weight matrix can be considered as a control variable for spatial autocorrelation. Hence, incorporating eigenvectors into regression models can help to reduce the effect of spatial autocorrelation, thus reduce model bias and enhancing model fitting (Griffith, 2003; Chun & Griffith, 2011).

To address these research gaps, this study aims to examine the spatial-temporal patterns of talent in China from 2000 to 2015 drawing on a unique four-year city-level panel data set built from population censuses and 1% population sample surveys. We apply a range of measures of inequality to assess the imbalance degree of the spatial distribution of highly educated talent in China. To tackle any estimation bias arising due to spatial autocorrelation (Griffith, 2003), we build an eigenvector spatial filtering negative binomial panel model (ESF-NBPM) to examine the determinants of the geographical distribution of talent. The contribution of this article is threefold. First, we analyzed the spatial patterns of talent in China at a finer geographical resolution (i.e., the city-level), compared to previous analysis based on provincial-level data. Our more detailed geographic analysis will contribute to the literature on the migration patterns of Chinese talent by developing an understanding of the intra-provincial differences in the locational choices of talent and providing more accurate identification of the key underpinning factors of these choices. Second, our analysis also contributes to our understanding of changes in the relative importance of local factors shaping the distribution of talent over time. Third, we applied the ESF specification to tackle spatial autocorrelation moving away from traditional gravity models which omit the spatial interconnectivity in the distribution of talent stocks.

The structure of this paper is as follows: Section 2 reviews the existing literature on spatial patterns and determinants of talent. Data sources and research methods are described in Section 3. Sections 4 determines the spatial concentration or evenness in the spatial distribution of talent in China and analyses its evolution over time based on a range of inequality measures. Section 5 constructs an ESF-NBPM to identify the key driving factors of the spatial distribution of talent. The final section summaries the key findings and discusses policy recommendations.

2. Literature review

2.1 Determinants of the geographic distribution of talent

Economic theories have made considerable efforts to identify the factors shaping the spatial distribution and redistribution of talent (Lewis, 1954; Sjaastad, 1962; Todaro, 1969). New classical theory assumes migration and redistribution as the resulting process of individuals thinking rationally and independently, aiming to maximize their utility (Greenwood, 1975). Empirical results evidence that differentials in regional economic opportunities are dominant factors influencing the spatial distribution of talent. The skilled labor force is spatially focused and thus reinforcing agglomeration economics (Greenwood, 1975; Harris & Todaro, 1970; Ranis & Fei, 1961). Recent research has shown that highly skilled and educated migrants tend to move to regions with high-income levels and abundant job opportunities (Arntz, 2010; Liu & Shen, 2014b; Rowe 2013a; Scott, 2010). Employment stability and job opportunity also play an important role in attracting skilled labor (Fielding, 1989; Findlay et al., 2008; Xu & Ouyang, 2018).

Alternatively, additional theories and models have been developed which emphasize the role that amenity differentials play in the redistribution of talent. They assume that spatial differences in economic opportunities reflect largely compensating differentials related to corresponding spatial differences in amenities (Graves 1976, 2018). Based on the principles of Graves' *equilibrium* model, Glaeser et al. (2001), Clark et al. (2002) and Florida (2002) expanded economic-driven migration models to include urban amenities. Empirical research has shown that strong relationship between regional amenities and redistribution of talent exists, particularly in terms of natural environmental amenities (Knapp & Gravest, 1989; Partridge, 2010; Rappaport, 2009), urban public services (Woodward et al., 2006; Rowe et al. 2017), and urban consumer facilities (Glaeser & Gottlieb, 2006).

The location of universities has emerged as a key urban amenity factor displaying a significant relationship with the distribution of highly educated individuals (Florida et al. 2008; Rowe 2013). Universities are key producers of talent. In less-attractive and underdeveloped regions, universities may also play the role of "talent exporters" since local highly educated graduates tend to move to other regions for employment (Florida et al., 2006). When talent is less mobile and has a higher settlement intention of the region, universities have a more notable effect on the pool of local talent (Qian, 2010).

2.2 Geography of talent in China

There has been a recent surge in the number of studies examining the spatial patterns of talent in China and their underpinning factors (Gu et al., 1999; Liu & Shen, 2014a, 2014b; Liu & Xu, 2017; Nie & Liu, 2018; Yu et al., 2019). Collectively, these studies revealed an unbalanced spatial distribution at city or provincial scales (Liu & Shen, 2014a, 2014b; Nie & Liu, 2018). They have also demonstrated the critical role of economic factors influencing the mobility decisions of a range of different types of talent labor, including highly skilled individuals (Miu et al., 2013), scientists (Zhang et al., 2011), and highly educated people (Liu & Shen, 2014a). They have shown that economic factors including GDP, average wage, unemployment rate, and industrial structure comprise the fundamental forces shaping the spatial distribution of China's population and talent (Liu & Shen, 2014b; Yu et al., 2019). They have also indicated that the uneven distribution of talent has been the result of a pronounced gap between developed and undeveloped regions in terms of economic development and living conditions (Liu & Shen, 2014a). Gu et al. (1999) revealed that economic factors, such as job opportunities and living conditions, are the main drivers of population distribution. Furthermore, the siphoning effects of talent spurring economic development in major cities may negatively impact small and medium-sized settlements, reinforcing spatial economic inequalities.

Consistent with Graves (1976, 2018) and Glaeser (2001), scholars have explored the role of amenities in influencing the distribution of talent in China (Deng et al., 2016; Liu & Shen, 2014b). Evidence indicates that public services, such as education and medical care represent a necessary condition to attract talent (Liu & Shen, 2014b). The local share of talent is also correlated with the local quality of the living and consumption facilities, including leisure and entertainment infrastructure, public safety, and shopping options (Deng et al., 2016). With the improvement of urban and inter-regional transportation infrastructure in China, regional traffic accessibility is also found to be positively associated with the distribution of talent (Gu et al., 2019a).

Recently, scholars have compared the impact of economic opportunities *versus* amenities on the mobility patterns of talent in China. The cumulative evidence thus far suggests that economic opportunities tend to prevail over amenity-driven migration (Gu et al., 2019a; Liu & Shen, 2014b; Yu et al., 2019). However, most prior empirical analyses have adopted a cross-sectional perspective on the distribution of talent, neglecting the temporal dynamics of this process. China has continued to experience sustained economic growth in the last decade with focused investment in eastern and south-eastern areas of the country (Breznitz & Murphree, 2011). These concentrated poles of economic development are likely to have enticed large flows of talent, generating losses elsewhere, Additionally, recently released population census and sample survey data have rarely been used to investigate the spatial patterns and determinants of talent in China over the last decade. The resources needed to obtain city-level Chinese talent data are costly since the data can only be derived from disaggregated census or 1% sample survey data books of provinces, usually published years after the census year, or the confidential 0.1% micro-level datasets, which are under highly restricted access.

2.3 Application of spatial modeling approach in population studies

Spatial autocorrelation is an inherent property in human migration and geographic data. Prior work has consistently shown the presence of strong and positive spatial autocorrelation in the spatial distribution of talent in China (Gu et al., 2019a). Six potential mechanisms that may lead to patterns of spatial autocorrelation have been identified. First, the spatial concentration of talent in an area has been found to lead to a reduction in communication cost of talent in neighboring regions (Fujita & Thisse, 1996). Second, it may also lead to an increase in talent volumes in neighboring regions, to take advantage of the existence of agglomeration economies (Miguélez et al., 2010). Third, regional economic growth trigger talent surge as a result of a knowledge spillover effect from adjacent cities with a large stock of talent and technology innovation enterprises (Henderson, 2007). Fourth, mobility of talent searching for jobs and residence tends to be spatially focused around their current region of residence and neighboring regions (Liu & Shen, 2017). Fifth, factors influencing the location of talent in a particular region spills over to nearby areas (Miguélez et al., 2010). Sixth, the inter-regional social network of talent reinforces concentration around local areas (Gu et al., 2019a). We recognize all these mechanisms producing spatial autocorrelation may result from the modifiable areal unit problem as often data used for spatial analysis are based on arbitrary administrative boundaries (MAUP: Chi & Zhu, 2008, Rowe 2017).

When modeling the process of the distribution of talent, traditional econometric models assume independent and identically distributed (i.i.d.) in the model residuals. Without controlling the effect of spatially structured random component (i.e., spatial autocorrelation), model residuals may contain significant spatial autocorrelation, thus violating the i.i.d. assumption and leading to endogeneity (Griffith, 2003; Gu et al., 2019b). Therefore, appropriate methods should be applied to reduce model bias resulting from spatial autocorrelation, such as the autoregressive model (Anselin, 1988) and the Getis filtering approach (Getis & Griffith, 2002). By adding the selected eigenvectors of the spatial weight matrix as the proxies for spatial autocorrelation, eigenvector spatial filtering (ESF) is another approach for tackling spatial autocorrelation in data. Comparing to other approaches for filtering spatial autocorrelation, ESF is easier to implement for various types of models and does not change the estimation method of the model. Thus it is more flexible and has fewer restrictions when capturing spatial autocorrelation (Griffith, 2003).

3. Research area, methods, and data

3.1 Areas

Our research area includes 31 provinces in China, excluding Hong Kong, Macao, and Taiwan. City extents comprise our geographic unit of analysis because they typically represent the spatial scale at which policies are formulated for the attraction and retention of talent, and factors influencing talent often vary significantly accross cities (Gu et al., 2019a; Nie & Liu, 2018). Based on individual-level records from population censuses and sample surveys, the stock of talent is aggregated to the city level. A total of 309 cities are chosen for our basic dataset. Cities in China include prefecture cities and autonomous cities. When we do regression analysis, we excluded data on autonomous cities because of their poor quality and number of missing entries. We used a total of 233 cities out of 309 units for regressions. China's economic geography is divided into four economic-geography regions: the Eastern, Central, Western, and Northeastern regions (Figure 1).

<Figure 1 about here>

3.2 Methods

Our analysis is divided into two stages. We first used a range of indicators of inequality to measure the extent of spatial concentration of talent before modeling their migration patterns in a negative binomial framework to identify the key underlying factors shaping these patterns, and ESP to account for spatial autocorrelation. Next, we described the indicators of inequality used.

Concentration Index (CI). The CI measures the spatial concentration of talent (Yang et al., 2014). This measure can be expressed in terms of the proportion (%) of talent in 1% of a country's land area. It is represented as:

$$CI_i = \frac{P_i/P_n \times 100\%}{A_i/A_n \times 100\%} = \frac{P_i/A_i}{P_n/A_n} \quad (1)$$

where CI_i denotes the CI of city i; P_i is the talent stock in city i; A_i is the land area of city i; A_n is the total land area of China; and, P_n is China's total stock of talent.

Coefficient of variation (CV). The CV complements the CI by measuring the degree of differences in the distribution of talent across the individual spatial unit and is calculated as follows (Haining & Haining, 2003):

$$C_{\text{v}} = \frac{1}{\bar{P}}\sqrt{\frac{\sum_{i=1}^{n}(P_i-\bar{P})^2}{n-1}} \quad (2)$$

where $\bar{p}$ denotes the average stock of talent across cities; n is the number of cities; and, p_i is the stock of talent in each city.

Lorenz Curve (LC). The LC graphically represents the degree of inequality of the distribution of talent across China's 309 cities. The LC was generated by plotting the percentage of highly educated population in each city against city layout in ascending order by their talent stocks. The mathematic definition of LC comes from Gastwirth (1971).

To model migration flows and identify the patterns underpinning their key drivers, we use a *negative binomial model (NBM*) framework which is appropriate to model count data (Cameron & Trivedi, 2013): our dependent variables are counts of talent i.e. number of people with a college degree or above. A basic regression modeling structure is the Poisson model (PM). However, the underlying assumption of the PM is equidispersion i.e. the variance of the dependent variable is assumed to be equal to the mean (Gu et al., 2019b; Liu & Shen, 2017) and this assumption is usually violated in migration analysis, yielding biased estimations (Rowe 2013b). To address this problem, the NBM introduces a parameter α that measures overdispersion in the data. For our dependent variables, the variance was significantly larger than the mean, so that we apply an individual-specific fixed effects NBM to model panel data taking the form of:

$$\Pr(\mathrm{Y} = y_{it}|\mu_{it}, \alpha) = \frac{\Gamma(\alpha^{-1}+y_{it})}{\Gamma(\alpha^{-1})+\Gamma(y_{it}+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_{it}}\right)^{\alpha^{-1}}\left(\frac{\mu_{it}}{\alpha^{-1}+\mu_{it}}\right)^{y_{it}} \quad (3)$$

$$\mu_{it} = \exp\left(\alpha_0 + \beta x_{it} + \varepsilon_i\right) \quad (4)$$

where Γ is the Gamma integral; μ_{it} equals $\mathrm{E}(y_{it})$; α is the variance parameter of the Gamma distribution. When this parameter tends to 0, the NBM becomes a PM; y_{it} represents the talent stock of city i in time t; x_{it} represents a vector of independent variables of city i in time t; α_0 is the constant term; β represents the vector of estimates capturing the relationship between the independent and dependent variables; and ε_i represents the error term.

To correct for spatial autocorrelation in our regression estimates, we use *Eigenvector spatial filtering (ESF).* The existence of spatial autocorrelation may lead to biased regression estimates, drawing erroneous conclusions and as a result in population migration and distribution data misguided policy recommendations (Griffith, 2003; Casado-Díaz et al. 2017). Eigenvectors of a given spatial weight matrix can represent components with different degrees of spatial autocorrelation. By adding these eigenvectors into regression models, the effect of spatially structured random components is controlled. Hence, ESF has a strong capacity to reduce spatial autocorrelation in residuals (Chun & Griffith, 2011; Fischer & Griffith, 2008; Griffith, 2003; Gu et al., 2019b; Liu & Shen, 2017).

In this study, ESF is implemented as follows: first, we constructed an n-by-n binary spatial weight matrix $\boldsymbol{W}$ under *queen* criteria[1] using GeoDa to represent the connectivity of Chinese cities (n denotes the total number of cites). The second step is to construct a transformed spatial weight matrix. Given a matrix $\mathbf{M} = (\mathbf{I} - \boldsymbol{A}\boldsymbol{A}'/n)$, where $\mathbf{I}$ is an n-by-n identity matrix, and $\boldsymbol{A}$ is an n-by-1 vector of 1s. $\mathbf{M}$ can center the spatial weight matrix $\boldsymbol{W}$ by $\mathbf{M}\boldsymbol{W}\mathbf{M}$. Third, we calculated the eigenvalues and eigenvectors of the matrix $\mathbf{M}\boldsymbol{W}\mathbf{M}$:

$$\mathbf{M}\boldsymbol{W}\mathbf{M} = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}' \tag{5}$$

where $\boldsymbol{E}$ is the matrix of eigenvectors decomposed by the transformed matrix, and $\boldsymbol{\Lambda}$ is a diagonal matrix with corresponding eigenvalues.

Getis and Griffith (2002) demonstrated that the eigenvectors in the ESF method are orthogonal and uncorrelated, and each eigenvector corresponds to a unique pattern of spatial autocorrelation. Additionally, each individual eigenvector in $\boldsymbol{E}$ can be linked to a Moran's *I*:

$$I_j = \frac{n}{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}} \frac{\boldsymbol{e}_j'\mathbf{M}\boldsymbol{W}\mathbf{M}\boldsymbol{e}_j}{\boldsymbol{e}_j'\mathbf{M}\boldsymbol{e}_j} \tag{6}$$

where n is the number of cities. I_j is the Moran's *I* for the j^{th} eigenvector $\boldsymbol{e}_j$. The eigenvector ($\boldsymbol{e}_1$) with the largest eigenvalue has the largest Moran's *I* (I_1) identifying the highest level of spatial autocorrelation in the dataset. The eigenvector ($\boldsymbol{e}_2$) with the second largest eigenvalue has the second-largest Moran's *I* (I_2) and indicates the second-highest spatial autocorrelation, and so on. In order to choose an appropriate set of eigenvectors controlling spatial autocorrelation in data, Griffith (2003) outlined that candidate eigenvectors can be identified with a critical value of the corresponding eigenvalues indicating a specific minimum spatial autocorrelation level (e.g., Moran's $I = 0.25$). After the selection of the candidate eigenvectors, a subset of eigenvectors can be chosen with the Akaike information criterion (AIC), and a linear combination of the selected eigenvectors can be used to capture latent spatial autocorrelation effects (Chun, 2011). This process minimizes the estimation error without reducing too much the degree of freedom of the model.

[1] The *queen* spatial weight mareix has been widely applied in defining the spatial relationship of China's cities (Gu et al., 2019a; Hong & Sun, 2011; Liu et al., 2019).

For space-time panel data analysis (e.g., the NBPM), eigenvectors need to be concatenated T times to match the total number of space-time observations (Chun, 2011). Because the spatial structure of the data is invariant over time, linear mixed models (LMM) and generalized linear mixed models (GLMM) are suggested to apply to account for spatial and temporal autocorrelation with ESF technique (Chun, 2011; 2014; Patuelli et al., 2011). Fortunately, the fixed effects NBPM specification proposed by Hausman et al. (1983) can accommodate the time-invariant spatial autocorrelation represented by the selected eigenvectors (Gu et al., 2019b). After adding eigenvectors that are used to control the effect of spatial autocorrelation, our model specification can be defined by:

$$\mu_{it} = \exp\left(\beta x_{it} + \gamma e_{it} + \varepsilon_i\right) \tag{7}$$

where e_{it} denotes the selected eigenvectors of city i in time t, and γ is the vector of estimators.

3.3 Data

Following existing studies (Gu et al., 2019a; Liu & Shen, 2014a; 2014b), we define talent: individuals with a college degree or above. The rationale of using academic qualifications to define talent is that academic qualifications provide an accurate representation of human capital embedded in the labor force (Schultz, 1961). In most cases, regional development policies of cities often use academic qualifications to define and divide talent (Rowe et al., 2013; Faggian et al. 2016).

We used data from the 2000 and 2010 Chinese population censuses and the 2005 and 2015 1% population sample surveys. In 2000, there were 45.6 million highly educated individuals in China, comprising 3.67% of the total population. By 2005, this number had increased by 30 million to comprise 75.7 million, and by 74 million to represent 120.1 million in 2010. In 2015, the pool of Chinese talent comprised 151.1 million and accounted for 10.99% of the total population.

For the independent variables, we used to explain the spatial distribution of talent in China. We draw on data from 1999, 2004, 2009, and 2014 reported in 2000, 2005, 2010, and 2015 *China City Statistical Yearbook* (National Bureau of Statistics, 2001, 2006, 2011, 2016). Including time-lagged independent variables in our models helps reducing potential endogeneity caused by reverse causation. Additionally, the use of ESF effectively reduces the effect of spatial autocorrelation in residuals and further alleviates any potential endogeneity (Getis & Griffith, 2002; Gu et al., 2019b). As in previous studies (Gu et al., 2019a; Liu & Shen, 2014a; 2014b), we included independent variables to capture two broad sets of factors: economic and amenity factors, and also introduced a range of control variables. A description of variables is provided in Table 1 for the full set of variables and descriptive statistics are reported in Table 2.

<Table 1 about here>

<Table 2 about here>

4. The spatial pattern of talent in China

This section determines the extent of spatial concentration of talent across cities in China. It first delves into the overall characteristics of spatial patterns of talent based on the CI, CV, and LC followed by an analysis of the spatial heterogeneity in these spatial patterns.

4.1 Concentrated pattern

Following Liu et al. (2010), we divided 309 cities into three types of areas based on their CI score: intensively-distributed areas (CI⩾2), evenly-distributed areas (0.5＜CI＜2), and sparsely-distributed areas (CI⩽0.5). We analyzed the distribution of talent across these areas and the results are reported in Table 3. We found that talent tended to concentrate on a small number of intensively-distributed areas with an increasing density from 32.28 persons/km^2 in 2000 to 99.97 in 2015. Just over 10% of China's land area clustered about 70% of the highly educated population during the 2000-2015 period.

<Table 3 about here>

Most of the highly educated population has tended to concentrate in large urban agglomerations and provincial capitals, particularly in the eastern coastal region of the country (Figure 2). This region encompasses intensively-distributed, economically developed areas, including first-tier cities such as Beijing and Shanghai, capital cities in the central and eastern regions such as Zhengzhou, Taiyuan, Wuhan, and Nanjing, and the cities in Pearl River Delta urban agglomerations such as Dongguan, Zhuhai, and Zhongshan.

<Figure 2 about here>

Evenly-distributed areas are mainly located in the central-eastern areas of China and have seen a fluctuation in the share of highly educated population, with a slight rise in 2015. Yet, these areas displayed a significant increase in the density of talent . Sparsely-distributed areas emerged persistently in western and north-west parts of the country. They occupy the majority of the land area (more than 60%) but has consistently contained only less than 6% of China's highly educated population. Though they recorded a small increase in density reflecting a reduction in land area.

4.2 Unbalanced distribution

We used the CV and LC to measure the degree of spatial inequality of talent. As shown in Figure 3, the CV returned a score of 1.576 in 2000 to 1.634 in 2010 and down to 1.472 in 2015, indicating a persistent pattern of high spatial concentration in the density of highly educated population across cities, with little variation over time. The increasing uneven trend in the first ten years is closely related to a significant difference in college enrollment expansion in particular cities, including first-tier cities (e.g., Beijing and Shanghai) and provincial capitals (e.g., Wuhan and Chengdu) after the introduction of the proposed *Action Plan for Education Revitalization in the 21st Century*. This plan has resulted in a variation in the local supply of educated people across cities. The slight decrease in the concentration of talent between 2010 and 2015 may reflect recent national initiatives of regional development to reduce spatial inequalities (e.g., *the National Medium- and Long-Term Talent Development Plan (2010-2020)* and *The National New-Type Urbanization Plan (2014-2020)*.

<Figure 3 about here>

Similarly, the LC also evidences an unbalanced distribution of talent in China (Figure 4). The vertical axis represents the cumulative percentage of talent stock, while the horizontal axis represents the cumulative percentage of cities ordered by talent stock in each city. The Equality Line (EL) shows a hypothetical balanced distribution of talent. The closer the distance between the LC and EL indicates an evener distribution of talent. The results show a closer LC to the EL for 2015 than for earlier years indicating that a slightly more even distribution of talent in 2015. Despite this decrease, the pattern of talent concentration is still strongly unbalanced and spatially concentrated: a large proportion of the talent (40-45%) concentrated in a small number of large cities (10%).

<Figure 4 about here>

4.3 Regional heterogeneity

Underpinning this national pattern of spatial concentration, stark regional differences in the distribution of talent exists. To explore this, we calculated the average density of talent in China's four economic-geography regions from 2000 to 2015 (Figure 5). The eastern region consistently recorded the highest density of talent, displaying a significant increase from 20 persons/km^2 in 2000 to over 70 in 2015. The western region reports the lowest and most marginal density level remaining below 10 people/km^2 since 2000.

<Figure 5 about here>

5. Determinants of talent distribution

This section explores the determinants of the spatial pattern of talent in Chinese cities based on the ESF-NPBM, considering the effect of spatial autocorrelation in the data.

5.1 Model processing

The dependent variable is the talent stock in each city (COLLEGE). Drawing on previous studies (Clark & Cosgrove, 1991; Clark & Hunter, 1992; Fielding, 1989; Palivos & Wang, 1996; Partridge, 2010), we defined 14 independent variables. Four variables were used to capture differences in economic opportunities across cities: gross domestic product (GDP), average annual wage (WAGE), unemployment rate (UNEMP) and industrial structure (INDUS). Four variables were used to capture differences in public sector amenities: including the ratio of total expenditure on science, technology and education (STEEXPEND), the ratio of fiscal expenditure to revenue (SPEND), the number of primary school teachers per 10,000 primary school students (PRIEDU) and the number of doctors per 10,000 people (MEDICAL). We also included three variables to account for differences in natural amenities: green coverage rate (GERRN), sewage treatment compliance rate (SEWAGE) and sulfur dioxide emissions (SO2). Additionally, we introduced three other control variables which may help to explain the distribution of talent: fixed-asset investment (FAI), population density (DENS) and the number of college students (UNISTU). To test the robustness of our models, we tested two dependent variables: the number of people with a college degree or above (COLLEGE), and the number of people with an undergraduate degree or above (UNDERGRA).

We tested for strict multi-collinearity of models with the VIF test under the assumption of ordinary least squares (OLS). The VIF values of each variable were all less than 4, indicating that there was no strict multi-collinearity problem. We also calculated the covariance matrix for all variables and found that the pair-wise correlation coefficients did not exceed 0.7. Considering the possible heteroscedasticity, we used a bootstrap technique (resampling 400 times) to compute cluster-robust standard errors (Chen, 2010).

To test for spatial autocorrelation, we calculated a panel-type standardized Moran's *I* for talent stock in 233 cities during 2000-2015 (Arbia & Piras, 2005). The Moran's *I* for this test was 0.125 ($p<0.01$), indicating statistically significant and positive spatial autocorrelation in the spatial pattern of talent in cities. To reduce the effect of the spatial autocorrelation, we employed the ESF specification to perform our regression analysis. The model results are reported in Table 4. Model 1 corresponds to a standard NBPM, including all 14 explanatory variables but no eigenvector, while for Model 2-6 we used a specification based on ESF and varying groupings of independent variables. The AIC is used to assess model fit. Comparing the AICs for Model 1 and Model 5 returns a much lower AIC for the latter indicating better model fit while controlling for spatial autocorrelation. We thus based our subsequent analyses on the results from the ESF NBPMs. Specifically, we analyzed: Model 2 including only economic-related variables; Model 3 including only amenity-related variables; Model 4 incorporating economic-related and amenity-related variables; Model 5 incorporating control variables; Model 6 based on a different dependent variable (i.e., COLLEGE).

5.2 Model results

Results from Model 2 showed statistically significant coefficients for GDP (GDP), average annual wage (WAGE), and unemployment rate (UNEMP). Results from Model 3 reveal significant coefficients for scientific technology and education expenditure (STEEXPEND), the number of primary school teachers (PRIEDU), and the number of doctors (MEDICAL). The greening rate (GREEN) and sewage treatment rate (SEWAGE) appear, however, insignificant, while sulfur dioxide emissions (SO2) shows a positive significant correlation to the local count of talent running against our expectation. This finding might be because when the effects of economic opportunities were not controlled, sulfur dioxide emissions may have become a good proxy for the local economic level reflecting the correlation between air quality and economic activity. Results from Model 4 show similar magnitude of coefficients for economic-related variables as in Model 2, while we observe changes for amenity-related factors, particularly MEDICAL and SO2 becoming statistically insignificant, and GREEN resulting in a significantly positive correlation.

<Table 4 about here>

Comparing Model 5 with Models 2-4 reveals that economic opportunities were the primary factors influencing the spatial distribution of talent in China between 2000 and 2015. GDP reflects higher economic benefits for local communities and population in terms of income and consumption spillover effects, generating further consumption in the local economy as well as job creation. The results indicate that an average of a 1% increase in urban GDP led to a 0.3202% increase in the stock of talent between 2000 and 2015. WAGE also exerted a significant impact on shaping the distribution of talent in China, with a 1% increase in the average annual wage of urban employees resulting in a rise in the local talent stock of 0.1848%. Unemployment also seems to have played a role. Selecting a workplace is often a risk-averse decision-making process for educated workers; thus our results indicate a negative correlation between the urban unemployment rate and the stock of talent. For every 1% increase in the unemployment rate, the stock of talent reduced by 0.0097%. The industrial structure of the local economy does not seem to play an important role in influencing the pool of local talent, which is partly due to the differences in industrial structure are small across most cities.

Specific urban amenities were also significantly associated with the spatial distribution of talent. The ratio of per capita science, technology, and education expenditure to financial expenditure (STEEXPEND) emerged as a key factor representing the importance placed by the local government on urban technology and education development. Our results indicate that a 1% increase in this ratio resulted in a rise in the stock of talent of the city of 0.0136%. In contemporary China, a large percentage of talent tends to migrate with their family and a primary concern is an education for their children (Gu et al., 2019c). The primary education context of destination cities is thus important in attracting talent. Our results show that if the number of primary school teachers per 10,000 students (PRIEDU) increased by 1%, the pool of talent expands by 0.2556%.

Other urban amenities also played a role but to a lesser extent. The rate of urban greening (GREEN) displays a positive relationship with talent stock. Model 5 suggests that a 1% increase in urban greening rate led to a 0.0023% increase in talent stock of the city. However, out of our expectation, other amenity variables were insignificant in our model, including the proportion of per capita fiscal expenditure to fiscal revenue (SPEND), sulfur dioxide emissions (SO2), the number of doctors per 10,000 (MEDICAL) and sewage treatment rate (SEWAGE), which implied that the relationship between urban amenities and the distribution of talent should be further studied. Particularly, after considering the effects of economic opportunities, GREEN becomes significant for model 4 and model 5. Also, the coefficients of SO2 and SEWAGE become positive, although they are insignificant. This indicates that economic opportunities are prerequisites for the influences of natural amenities. Besides, the effect of MEDICAL becomes insignificant after controlling economic-related variables, which is partly because the number of doctors per 10,000 is closely related to a regional economic level.

5.3 Robustness Checks

We observed the changes in the significance of individual covariates from Model 2 to Model 5 and found that a persistent pattern in the statistical significance and magnitude of economic variables. However, much more variability is observed for amenity-related variables. For instance, coefficients for sulfur dioxide emissions (SO2) and the number of doctors (MEDICAL) become insignificant after controlling for urban economic opportunities.

To test the robustness of these results, we used a different proxy for talent i.e. the number of people with a bachelor's degree or above (UNDERGRA) as a dependent variable and used the same specification as Model 5. Model 6 in Table 4 reports the results. Consistent with previous models, they show significant coefficients and similar signs for economic factors: GDP, WAGE and UNEMP. Changes are observed for amenity-related variables: GREEN is no longer significant, yet SEWAGE and SO2 become significant. This reflects the slight difference between the two types of talent's preferences for amenities. The urban greening rate had a closer relationship with the stock of talent with a college degree or above, while the sewage treatment compliance rate and sulfur dioxide emissions were more associated with the stock of talent with a bachelor's degree or above.

Overall, based on the results of the ESF NBPMs, urban economic opportunities appear to be essential forces shaping the spatial distribution of talent across Chinese cities. The evidence reveals that only certain amenity-related factors can be consistently be associated with the location of talent. Amenities related to the local provision of primary education, and investment in science, technology, and education seem to be critical.

6. Conclusion and Discussion

This paper assesses and seeks to identify key factors shaping the spatial distribution of talent across Chinese cities between 2000 and 2015. Results revealed a high degree of spatial concentration in the distribution of Chinese talent throughout the fifteen-year period in a small number of cities. While this concentration decreased between 2010 and 2015, it remains relatively high. Examining the density of talent within each of four economic-geography regions of China reveals a pattern of rapidly increasing concentration of talent in the eastern region, increasing the gap in the local stock of spatial talent between regions. Only small changes in the local pool of talent were observed in western and north-western cities.

A significant, positive spatial autocorrelation in the spatial distribution of talent across Chinese cities exists. An ESF specification was used to mitigate the effects of spatial autocorrelation and reduce estimation bias. Our modeling results revealed that urban economic opportunities have played a dominant role in shaping the spatial distribution of talent in China. Economic variables were consistently and significantly associated with the local pool of talent. Specifically, regional GDP and average annual wage of urban employees showed strong positive relationships with talent stock, while the urban unemployment rate was negatively correlated. This finding is consistent with the evidence from previous cross-sectional studies examining on the spatial distribution of China's skilled workers or highly educated people and their determinants, largely prior to 2010 (e.g., Liu & Shen, 2014a). Our work expands previous work by revealing that the systematic key role of economic factors in shaping migration patterns of talent in China over time.

Urban amenities seem to have played a less important role. Only specific variables consistently displayed statistically significant associations. Urban per capita science, technology and education expenditure ratio and the number of teachers per 10,000 primary school students showed a persistent relationship with the stock of talent in Chinses cities.

As China's economic resources and public service supplies are concentrated in developed metropolitan areas with large populations, the existing pattern of spatial concentration and unbalanced distribution in the pool of talent is expected to persist over time. Our results have wide-ranging policy implications. From a national and regional perspective, they emphasize the need for appropriate regional development policies to optimize the pattern of talent. For areas with a high density of talent, policies should be targeted at taking advantage of the positive externalities brought about by the clustering of talent. For the central and western regions, where talent is sparsely distributed, preferential policies should be formulated to give specific financial supports to people, such as young graduates to guarantee their fundamental lives. From a city's perspective, local governments should enhance the economic level of cities by offering higher wages and enhancing employment stability to attract talent. Meanwhile, governments should also focus on improving public services and the urban environment, particularly primary education amenities and raising local expenditure on science, technology and education, so that the city can meet the increasing and diverse needs of talent.

Using the data of Chinese highly educated talent at the city level, the present paper measures the average effects of economic- and amenity-related factors on the spatial distribution of talent over 2000-2015. China has gone through significant social changes over this period, which may have resulted in a varying effect of the factors shaping the geographic distribution of talent over time. Future studies may estimate the time-variant influence of these factors and how they vary across cities.

References

Abel, J. R., & Deitz, R. (2011). Do colleges and universities increase their region's human capital?. *Journal of Economic Geography*, *12*(3), 667-691.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer

Arbia, G., & Piras, G. (2005). Convergence in per-capita GDP across European regions using panel data models extended to spatial autocorrelation effects.

Arntz, M. (2010). What attracts human capital? Understanding the skill composition of interregional job matches in Germany. *Regional Studies*, *44*(4), 423-441.

Breznitz, D., & Murphree, M. (2011). *Run of the red queen: Government, innovation, globalization, and economic growth in China*. Yale University Press.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.

Casado-Díaz, J.M., Martínez-Bernabéu, L. and Rowe, F., 2017. An evolutionary approach to the delimitation of labour market areas: an empirical application for Chile. Spatial economic analysis, 12(4), pp.379-403.

Chen, Q. (2010). *Advanced econometrics and Stata applications.* Beijing: Higher Education Press.

Chun, Y., & Griffith, D. A. (2011). Modeling network autocorrelation in space–time migration flow data: an eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, *101*(3), 523-536.

Chun, Y. (2014). Analyzing space–time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. *Geographical Analysis*, *46*(2), 165-184.

Chi, G., & Zhu, J. (2008). Spatial regression models for demographic analysis. *Population Research and Policy Review*, *27*(1), 17-42.

Clark, T. N., Lloyd, R., Wong, K. K., & Jain, P. (2002). Amenities drive urban growth. *Journal of urban affairs*, *24*(5), 493-515.

Clark, D. E., & Cosgrove, J. C. (1991). Amenities versus labor market opportunities: Choosing the optimal distance to move. *Journal of Regional Science*, *31*(3), 311-328.

Clark, D. E., & Hunter, W. J. (1992). The impact of economic opportunity, amenities and fiscal factors on age-specific migration rates. *Journal of Regional Science*, *32*(3), 349-365.

Cowell, F. (2011). *Measuring inequality*. Oxford University Press.

Deng, J., Yin, T., & Zhou, L. (2001). Research on the Strategies and Determinants of the development of the Talents in Guangzhou. *South China Population, 2001*(4), 29-33.

Faggian, A., Corcoran, J. and Rowe, F., 2016. Evaluating the effects of Australian policy changes on human capital: the role of a graduate visa scheme. *Environment and Planning C: Government and Policy*, *34*(1), pp.151-170.

Florida, R. (2002). The economic geography of talent. *Annals of the Association of American Geographers*, *92*(4), 743-755.

Florida, R., Mellander, C., & Qian, H. (2012). China's development disconnect. *Environment and Planning A*, *44*(3), 628-648.

Florida, R. (2006). Where the brains are. *The Atlantic Monthly*, *298*(3), 34.

Florida, R., Mellander, C., & Qian, H. (2008). *Creative China? The university, tolerance and talent in Chinese regional development* (No. 145). Royal Institute of Technology, CESIS-Centre of Excellence for Science and Innovation Studies.

Fielding, A. J. (1989). Inter-regional migration and social change: a study of South East England based upon data from the Longitudinal Study. *Transactions of the Institute of British Geographers*, 24-36.

Findlay, A., Mason, C., Harrison, R., Houston, D., & McCollum, D. (2008). Getting off the escalator? A study of Scots out-migration from a global city region. *Environment and Planning A*, *40*(9), 2169-2185.

Fischer, M. M., & Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union. *Journal of Regional Science*, *48*(5), 969-989.

Fujita, M., & Thisse, J. F. (1996). Economics of agglomeration. *Journal of the Japanese and international economies*, *10*(4), 339-378.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica: Journal of the Econometric Society*, 1037-1039.

Getis, A., & Griffith, D. A. (2002). Comparative spatial filtering in regression analysis. *Geographical analysis*, *34*(2), 130-140.

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of economic geography*, *1*(1), 27-50.

Glaeser, E. L., & Gottlieb, J. D. (2006). Urban resurgence and the consumer city. *Urban studies*, *43*(8), 1275-1299.

Graves, P. E. (1976). A reexamination of migration, economic opportunity, and the quality of life. *Journal of Regional Science*, *16*(1), 107-112.

Graves, P. E. (2018). Spatial equilibrium in labor markets. *Handbook of Regional Science*, 1-17.

Greenwood, M. J. (1975). Research on internal migration in the united states: a survey. *Journal of Economic Literature,13*(2), 397-433.

Griffith, D. A. (1988). *Advanced Spatial Statistics: Special Topics in the Exploration* ***of*** *Quantitative Spatial Data Series.* Dordrecht: Kluwer

Griffith, D. A. (2003). *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer Science & Business Media.

Gu, H., Meng, X., Shen, T., & Wen, L. (2019a). China's Highly Educated Talents in 2015: Patterns, Determinants and Spatial Spillover Effects. *Applied Spatial Analysis and Policy*, 1-18.

Gu, H., Liu, Z., Shen, T., & Meng, X. (2019b). Modelling interprovincial migration in China from 1995 to 2015 based on an eigenvector spatial filtering negative binomial model. *Population, Space and Place*. e2253. https://doi.org/10.1002/psp.2253

Gu, H., Liu, Z., & Shen, T. (2019c). Spatial pattern and determinants of migrant workers' interprovincial hukou transfer intention in China: Evidence from a National Migrant Population Dynamic Monitoring Survey in 2016. *Population Space Place.* e2250. https://doi.org/10.1002/psp.2250

Gu, C., Cai, J., Zhang, W., Ma, Q., Chan, R. C. K., Li, W., & Shen, D. (1999). A study on the patterns of migration in Chinese large and medium cities. *Acta Geographica Sinica, 1999*(3), 14-22.

Haining, R. P., & Haining, R. (2003). *Spatial data analysis: theory and practice*. Cambridge University Press.

Harris, J. R. , & Todaro, M. P. (1970). Migration, unemployment & development: a two-sector analysis. *American Economic Review, 60*(1), 126-142.

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica, 52*(4), 909-938.

Henderson, J. V. (2007). Understanding knowledge spillovers. *Regional Science and Urban Economics*, *37*(4), 497-508.

Hong, E., & Sun, L. (2011). Foreign direct investment and total factor productivity in China: A spatial dynamic panel analysis. *Oxford Bulletin of Economics and Statistics, 73*(6), 771-791.

Kerr, S. P., Kerr, W., Özden, Ç., & Parsons, C. (2017). High-skilled migration and agglomeration. *Annual Review of Economics*, *9*, 201-234.

Kerr, S. P., Kerr, W. R., & Lincoln, W. F. (2015). Skilled immigration and the employment structures of US firms. *Journal of Labor Economics*, *33*(S1), S147-S186.

Kerr, W. R. (2013). *US high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence*(No. w19377). National Bureau of Economic Research.

Knapp, T. A., & Gravest, P. E. (1989). On the role of amenities in models of migration and regional development. *Journal of regional science*, *29*(1), 71-87.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The manchester school*, *22*(2), 139-191.

Liu, R., Feng, Z., & You, Z. (2010). Research on the spatial pattern & formation mechanisms of population agglomeration & shrinking in China. *China Population, Resources and Environment*, 20(3), 89-94.

Liu, Y., & Shen, J. (2014a). Spatial patterns and determinants of skilled internal migration in China, 2000–2005. *Papers in Regional Science*, *93*(4), 749-771.

Liu, Y., & Shen, J. (2014b). Jobs or amenities? Location choices of interprovincial skilled migrants in China, 2000–2005. *Population, Space and Place*, *20*(7), 592-605.

Liu, Y., & Shen, J. (2017). Modelling skilled and less-skilled interregional migrations in China, 2000–2005. *Population, Space and Place*, *23*(4), e2027.

Liu, Y., & Xu, W. (2017). Destination choices of permanent and temporary migrants in China, 1985–2005. *Population, Space and Place*, *23*(1), e1963.

Liu, Y., Shen, J., Xu, W., & Wang, G. (2017). From school to university to work: migration of highly educated youths in China. *The Annals of Regional Science, 59*(3), 651-676.

Liu, Y., Wang, R., Xue, D., & Zeng, J. (2019) The spatial pattern and determinants of skilled laborers and lessskilled laborers in China: Evidence from 2000 and 2010 censuses. *Geographical Research, 38*(8), 1949-1964.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, *22*(1), 3-42.

Yang, Y., Feng, Z., Wang, L., & You, Z. (2014). Research on the suitability of population distribution at the provincial scale in China. *Journal of Geographical Sciences*, *24*(5), 889-906.

McGranahan, D. A. (1999). *Natural amenities drive rural population change* (No. 1473-2016-120765).

National bureau of statistics. (2001). *China City Statistical Yearbook 2000.* Beijing: China Statistics Press.

National bureau of statistics. (2006). *China City Statistical Yearbook 2005.* Beijing: China Statistics Press.

National bureau of statistics. (2011). *China City Statistical Yearbook 2010.* Beijing: China Statistics Press.

National bureau of statistics. (2016). *China City Statistical Yearbook 2015.* Beijing: China Statistics Press.

Miguélez, E., Moreno, R., & Suriñach, J. (2010). Inventors on the move: Tracing inventors' mobility and its spatial distribution. *Papers in Regional Science*, *89*(2), 251-274.

Miu, G., Chen, W., & Tang, C. (2013) Study on Influence Factos of Talent Agglomeration in High-tech Enterprises——Taking Commercial Aircraft Corporation of China Ltd for Example. *Science and Technology Management Research, 33*(10), 120-122.

Nie, J., & Liu, H. (2018). Spatial Pattern and the Resulting Characteristics of Talent Flows in China. *Scientia Geographica Sinica, 38*(12), 1979-1987.

Palivos, T., & Wang, P. (1996). Spatial agglomeration and endogenous growth. *Regional Science and Urban Economics*, *26*(6), 645-669.

Partridge, M. D. (2010). The duelling models: NEG vs amenity migration in explaining US engines of growth. *Papers in Regional Science*, *89*(3), 513-536.

Patuelli, R., Griffith, D. A., Tiefelsdorf, M., & Nijkamp, P. (2011). Spatial filtering and eigenvector stability: space-time models for German unemployment data. *International Regional Science Review*, *34*(2), 253-280.

Qian, H. (2010). Talent, creativity and regional economic performance: the case of China. *The Annals of Regional Science*, *45*(1), 133-156.

Ranis, G., & Fei, J. C. H. (1961). A theory of economic development. *American Economic Review, 51*(4), 533-565.

Rappaport, J. (2009). Moving to nice weather. In *Environmental Amenities and Regional Economic Development* (pp. 25-53). Routledge.

Rodríguez-Pose, A., & Tselios, V. (2011). Mapping the European regional educational distribution. *European Urban and Regional Studies*, *18*(4), 358-374.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, *98*(5, Part 2), S71-S102.

Rowe, F., 2013a. The geography and determinants of regional human capital in eight Latin American and Caribbean countries. In Regional Problems and Policies in Latin America (pp. 379-405). Springer, Berlin, Heidelberg.

Rowe, F. 2013b. Spatial Labour Mobility in a Transition Economy: Migration and Commuting in Chile. Ph.D. thesis, University of Queensland.

Rowe, F, 2017. The CHilean Internal Migration (CHIM) database: Temporally consistent spatial data for the analysis of human mobility. *Region*, *4*(3), pp.1-6.

Rowe, F., Corcoran, J. and Faggian, A., 2013. Mobility patterns of overseas human capital in Australia: the role of a 'new'graduate visa scheme and rural development policy. Australian Geographer, 44(2), pp.177-195.

Rowe, F., Corcoran, J. and Bell, M., 2017. The returns to migration and human capital accumulation pathways: non-metropolitan youth in the school-to-work transition. *The Annals of Regional Science*, *59*(3), pp.819-845.

Todaro, M. P. (1969). A model of labor migration and urban unemployment in less developed countries. *The American Economic Review*, *59*(1), 138-148.

Scott, A. J. (2010). Jobs or amenities? Destination choices of migrant engineers in the USA. *Papers in Regional Science*, *89*(1), 43-63.

Schultz, T. W. (1961). Investment in human capital. *The American economic review*, 1-17.

Shen Municipal Government. (2019). *Opinions on the implementation of the "Peacock Plan" for the introduction of high-level overseas talents.* http://www.gd.gov.cn/zwgk/zcfgk/content/post_2531436.htmlSjaastad, L. A. (1962). The costs and returns of human migration. *Journal of political Economy*, *70*(5, Part 2), 80-93.

Song, H., Zhang, M., & Wang, R. (2016). Amenities and spatial talent distribution: evidence from the Chinese IT industry. *Cambridge Journal of Regions, Economy and Society*, *9*(3), 517-533.

State Council of China. (2010). *The National Medium and Long-Term Talent Development Plan (2010-2020).* http://www.gov.cn/jrzg/2010-06/06/content_1621777.htm.

Woodward, D., Figueiredo, O., & Guimaraes, P. (2006). Beyond the Silicon Valley: University R&D and high-technology location. *Journal of Urban Economics*, *60*(1), 15-32.

Xu, Z., & Ouyang, A. (2018). The Factors Influencing China's population distribution and spatial heterogeneity: a prefectural-level analysis using geographically weighted regression. *Applied Spatial Analysis and Policy*, *11*(3), 465-480.

Yu, Z., Zhang, H., Tao, Z., & Liang, J. (2019). Amenities, economic opportunities and patterns of migration at the city level in China. *Asian and Pacific Migration Journal*, 0117196819832054.

Zhang, J., Du, D., & Jiang, H. (2011) Regional Differences of S&T Talents in Jiangsu Province. *Scientia Geographica Sinica, 31*(3), 378-384.

Zhou, L., Tian, L., Gao, Y., Ling, Y., Fan, C., Hou, D., ... & Zhou, W. (2019). How did industrial land supply respond to transitions in state strategy? An analysis of prefecture-level cities in China from 2007 to 2016. *Land Use Policy*, *87*, 104009.

Zeng, J., Liu, Y., Wang, R., & Zhan, P. (2019). Absorptive Capacity and Regional Innovation in China: an Analysis of Patent Applications, 2000–2015. *Applied Spatial Analysis and Policy*, 1-19.

Table 1. Concentration pattern of talent from 2000 to 2015

	Talent proportion (%)	Land proportion (%)	Density (persons/km^2)	Density per year (persons/km^2)
2000				
Intensively-distributed area	66.96%	11.60%	32.38	
Evenly-distributed area	28.03%	26.36%	5.97	5.61
Sparsely-distributed area	5.01%	62.03%	0.45	
2005				
Intensively-distributed area	66.59%	12.09%	46.03	
Evenly-distributed area	27.16%	25.94%	8.75	8.36
Sparsely-distributed area	6.25%	61.97%	0.84	
2010				
Intensively-distributed area	70.65%	12.30%	85.09	
Evenly-distributed area	23.40%	23.64%	14.67	14.82
Sparsely-distributed area	5.28%	64.06%	1.22	
2015				
Intensively-distributed area	65.16%	11.67%	99.97	
Evenly-distributed area	28.84%	27.94%	18.48	17.90
Sparsely-distributed area	6.00%	60.39%	1.78	

Table 2. Description and expected effects of variables

Variable	Description	Expected effect
Dependent variables		
COLLEGE	Number of people with a college degree or above of each city in 2000, 2005, 2010, 2015	
UNDERGRA	Number of people with a bachelor degree and above of each city in 2000, 2005, 2010, 2015	
Economic opportunity variables		
GDP	Gross GDP of each city in 1999, 2004, 2009 and 2014 (10,000 yuan)	+
WAGE	The average annual wage of employees in urban areas of each city in 1999, 2004, 2009 and 2014 (yuan)	+
UNEMP	The urban unemployment rate of each city in 1999, 2004, 2009 and 2014 (%)	-
INDUS	The proportion of tertiary industry to GDP of each city in 1999, 2004, 2009 and 2014 (%)	+
Amenity variables		
STEEXPEND	The proportion of per capita science, technology and education expenditure to the financial expenditure of each city in 1999, 2004, 2009 and 2014 (%)	+
SPEND	The ratio of per capita financial expenditure to per capita fiscal revenue of each city in 1999, 2004, 2009 and 2014 (%)	+

PRIEDU	Number of primary school teachers per 10,000 primary school students of each city in 1999, 2004, 2009 and 2014	+
GREEN	Greening rate of each city in 1999, 2004, 2009 and 2014 (%)	+
SEWAGE	Sewage treatment compliance rate of each city in 1999, 2004, 2009 and 2014 (%)	+
SO2	Emissions of industrial sulfur dioxide of each city in 1999, 2004, 2009 and 2014 (tons)	-
MEDICAL	Number of doctors per 10,000 people of each city in 1999, 2004, 2009 and 2014	+
Control variables		
DENS	Population density of each city in 1999, 2004, 2009 and 2014 (persons/km^2)	+
UNISTU	Number of college students per 10,000 people of each city in 1999, 2004, 2009, and 2014	+
FAI	Per capita fixed assets investment of each city in 1999, 2004, 2009 and 2014 (10,000 yuan)	NS

Note: "+" denotes a positive expected effect, "-" denotes a negative expected effect, "NS" represents an unsure expected effect

Table 3. Descriptive statistics of variables

Variable	Number	Mean	Standard deviation	Minimum	Maximum
Dependent variables					
COLLEGE	932	359157	578240.7	11377	7241471
UNDERGRA	932	158344.1	413589.2	2018	6716225
Economic variables					
GDP*	932	12817129.2	21305227.4	168848	235677000
WAGE*	932	24754.39	17273.35	4189	103400.4
UNEMP	928	4.03	2.57	0.05	31.58
INDUS	931	36.49	8.39	12.1	77.95
Amenity variables					
STEEXPEND	932	18.13	5.32	2.53	37.36
SPEND	932	213.62	125.06	60.44	1802.50
PRIEDU *	931	559.47	148.53	189.93	1486.94
GREEN	929	35.27	13.90	0.7	92.87
SEWAGE	917	84.32	17.69	0.6	100
SO2*	924	47480.04	27879.18	0.1	641088
MEDICAL *	932	18.65	9.48	2.76	87.51
Control variables					
DENS	916	137.95	207.72	0	1293.38
UNISTU*	932	946.23	855.04	12	6161
FAI*	931	17494.63	20967.58	180.61	153802.5

Note: “” indicates logarithmic processing*

Table 4. Results from the NBPMs and ESF NBPMs

	(1) COLLEGE	(2) COLLEGE	(3) COLLEGE	(4) COLLEGE	(5) COLLEGE	(6) UNDERGRA
GDP	0.3509*** (0.0460)	0.2769*** (0.0397)		0.3291*** (0.0458)	0.3202*** (0.0510)	0.1108* (0.0629)
WAGE	0.1611** (0.0643)	0.3191*** (0.0411)		0.1986*** (0.0633)	0.1848*** (0.0695)	0.6318**** (0.0856)
UNEMP	-0.0101* (0.0055)	-0.0114** (0.0051)		-0.0095* (0.0054)	-0.0097** (0.0030)	-0.0194** (0.0090)
INDUS	0.0022 (0.0026)	0.0018 (0.0020)		0.0015 (0.0032)	0.0021 (0.0030)	-0.0020 (0.0030)
STEEXPEND	0.0140*** (0.0028)		0.0204** (0.0046)	0.0136*** (0.0029)	0.0136*** (0.0032)	0.0141*** (0.0056)
SPEND	0.0002 (0.0002)		0.0002 (0.0002)	0.0002 (0.0002)	0.0001 (0.002)	0.0007*** (0.0002)
PRIEDU	0.2800*** (0.0742)		0.4699*** (0.1402)	0.2573*** (0.0753)	0.2556*** (0.0798)	0.4307*** (0.1055)
GREEN	0.0024** (0.0011)		0.0034 (0.0060)	0.0024* (0.0013)	0.0023* (0.0011)	0.0005 (0.0011)
SEWAGE	-0.0006 (0.0007)		-0.0005 (0.0010)	0.0001 (0.0006)	0.0001 (0.0008)	0.0026** (0.0011)
SO2	-0.0015 (0.0054)		0.0558*** (0.0047)	-0.0049 (0.0044)	-0.0052 (0.0048)	-0.0324*** (0.0117)
MEDICAL	-0.0015 (0.0422)		0.8173*** (0.0724)	-0.0013 (0.0543)	-0.0168 (0.0633)	0.0533 (0.0797)
UNISTU	0.0122** (0.0052)				0.0142** (0.0062)	0.0291*** (0.0099)
FAI	-0.0081 (0.0249)				0.0016 (0.0301)	-0.0293 (0.0440)
DENS	0.0385* (0.0213)				0.0272 (0.0221)	0.1339*** (0.0415)
CONSTANT	-6.5960*** (0.5038)	-4.8593*** (0.2953)	-4.2884*** (0.8390)	-6.3053*** (0.4238)	-6.2126*** (0. 6132)	- 9.8892*** (0. 7632)
N	892	927	913	909	892	892
Eigenvectors	no	yes	yes	yes	yes	yes
AIC	15903.5	16762.9	16900.38	16274.3	15885.85	14988.07
Log likelihood	-7936.7494	-8361.4506	-8427.1896	-8110.1501	-7910.4024	-7464.0368

*Note: *** represents p＜0.01， ** represents p＜0.05， * represents p＜0.1; Cluster-robust standard errors are in parentheses.*

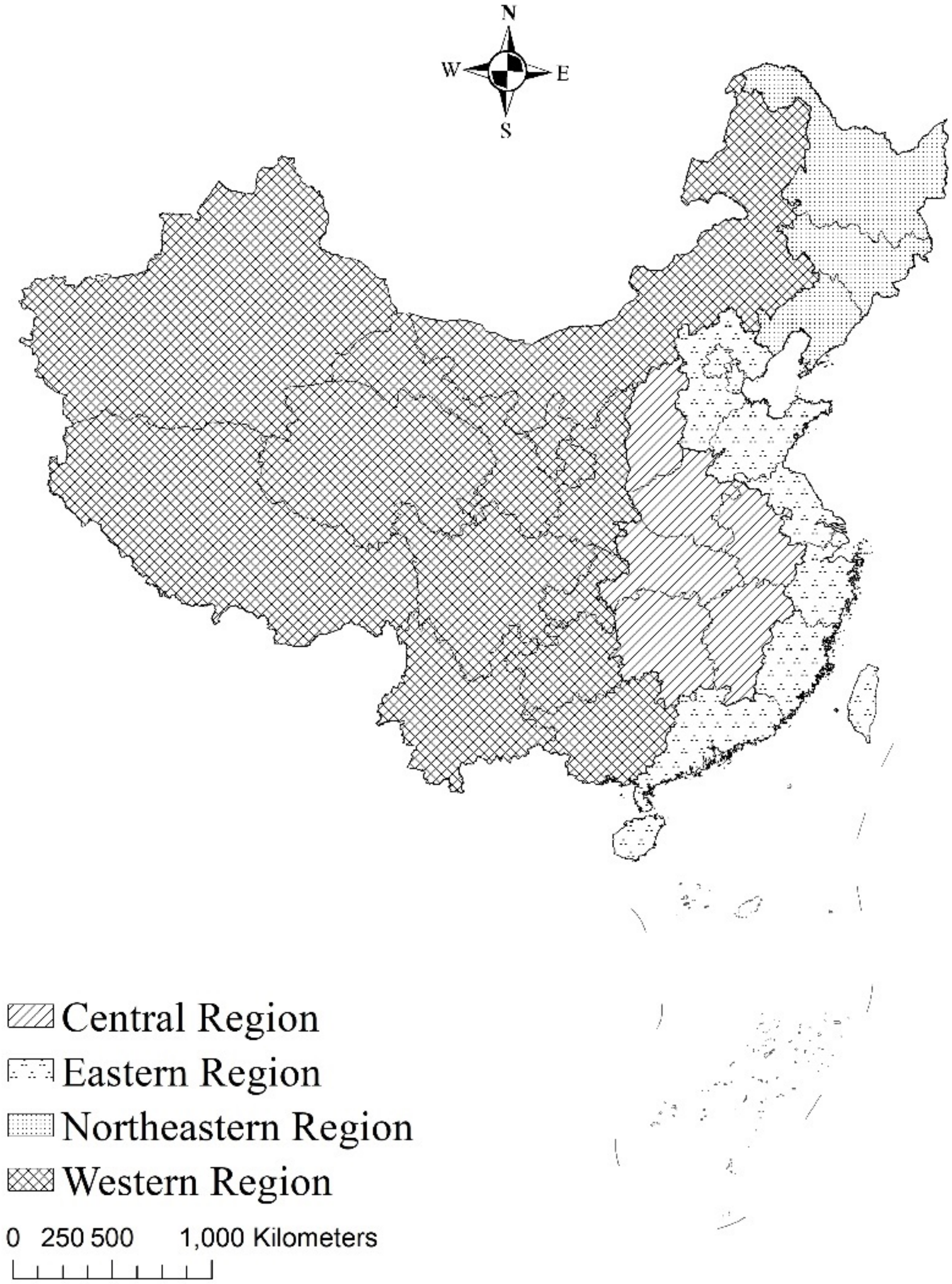

Figure 1. Economic regions in China

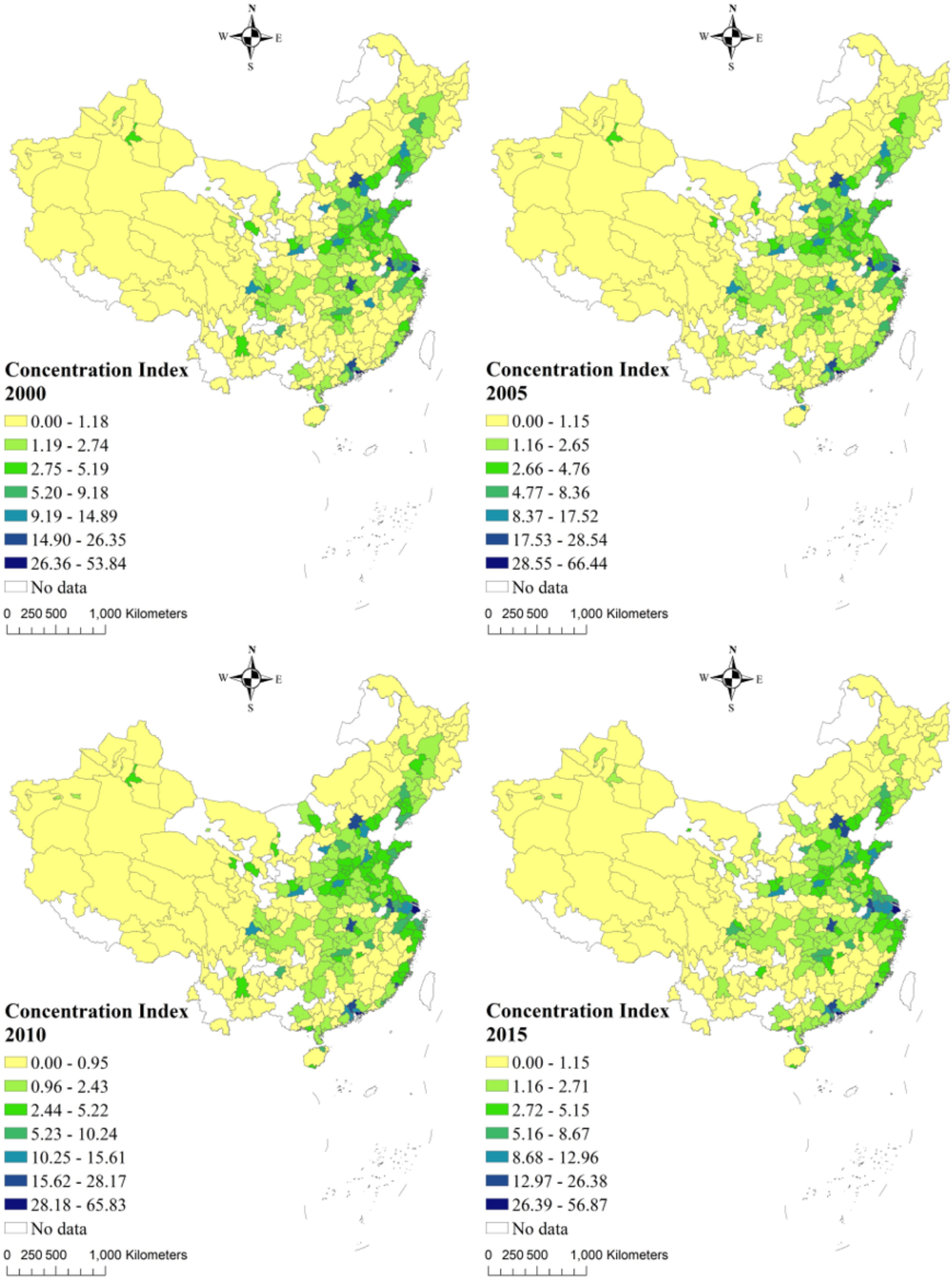

Figure 2. The spatial pattern of talent at the city-level, 2000-2015

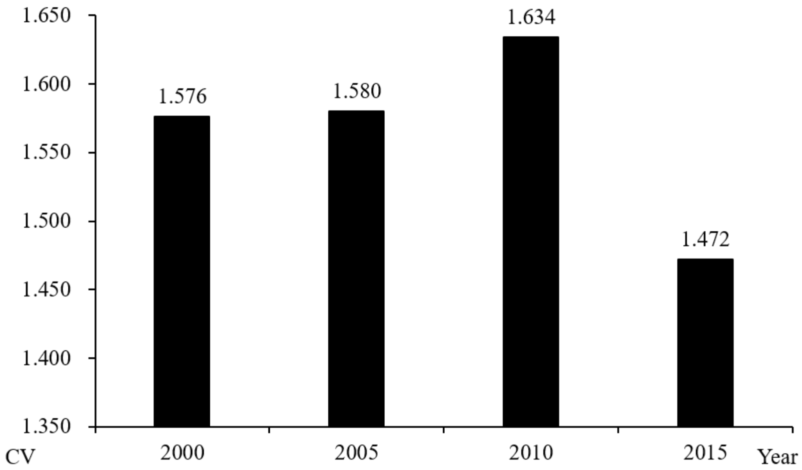

Figure 3. Coefficients of variation of the distribution of talent between 2000 and 2015

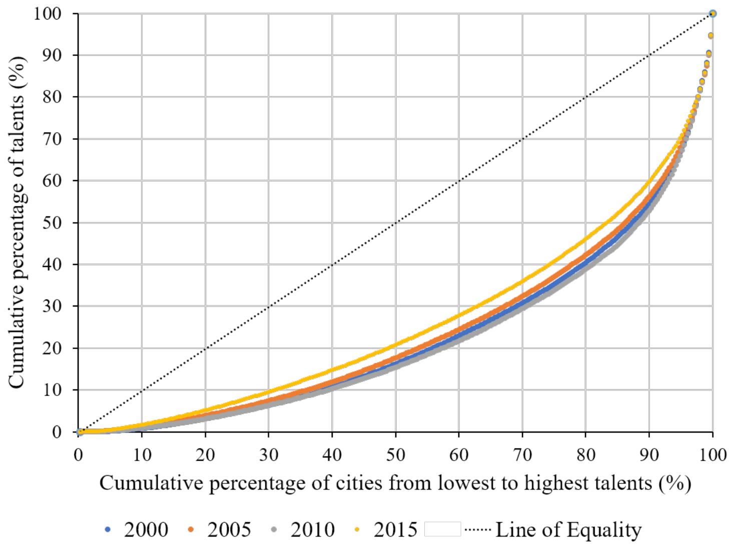

Figure 4. Lorenz curves of the stock of talent in each city, 2000-2015

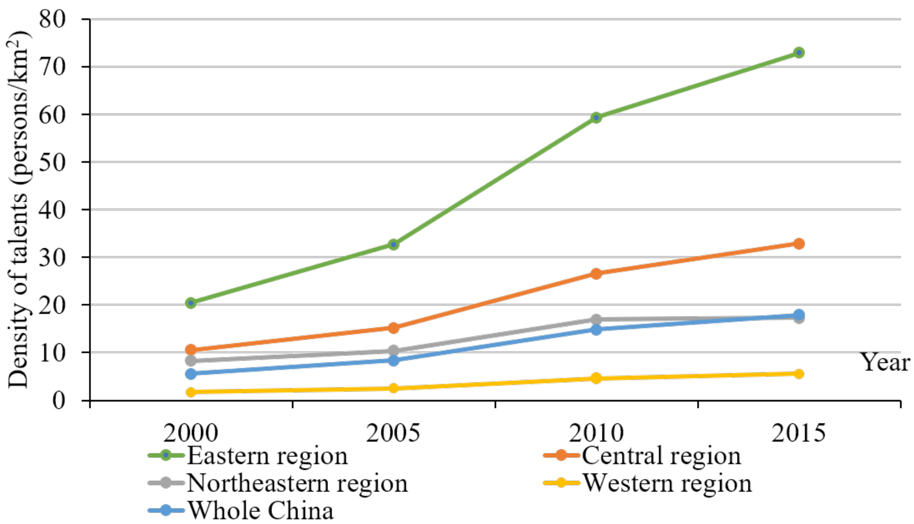

Figure 5. The density of talent in each economic-geography region and the whole of China from 2000 to 2015