Postprint: Spatial Econometric Analysis of Factors Influencing Population Distribution in Shaanxi Province

Authors: Mi Ruihua, Gao Xiangdong

Date: 2020-06-02T00:00:00+00:00

Abstract

Research on the influencing factors of population distribution facilitates the revelation of population distribution patterns and the prediction of population distribution trends. Based on county-level data from Shaanxi Province encompassing population, economic and social, and physical geography dimensions, this study employs factor analysis methods and spatial econometric modeling to dissect the influencing factors of population distribution. The findings indicate that the population areal differentiation ratio depends not only on observable external characteristics within a given district or county, including economic and social conditions, historical foundations, and physical geography, but also on unobservable, model-omitted common characteristics of that district or county, wherein the economic and public service factor and the population base factor exert the most significant positive explanatory power on population distribution, while the effects of other factors are relatively weak or statistically insignificant; urban hierarchy can significantly amplify the effects of industrial structure, per capita income, and topographic factors on population distribution. The study posits that economic and public service factors are pivotal for optimizing population distribution, while simultaneously necessitating consideration of the constraining effects of physical geography factors. This research offers valuable reference for policy formulation aimed at population distribution optimization.

Full Text

Preamble

This research was supported by the ChinaXiv cooperation journal program under grants 18ARK001, 19ARK004, 18JK0850, and 19BDD08. The authors declare no competing interests.

Author Information: JKLM (born 1983), female, Ph.D. candidate. Email: rice221@163.com. NOJK (born 1963), male, Ph.D. Email: xdgao@ecnu.edu.cn.

Abstract: Understanding population distribution and its evolutionary trends is essential for empirical research on the factors influencing population distribution. This study examines all 107 counties in Shaanxi Province, northwest China, using 2015 demographic, socioeconomic, and natural geographic data. We employ ordinary least squares (OLS) regression and spatial econometric models to analyze factors influencing Shaanxi' s population distribution. The demographic and socioeconomic data were extracted from the 2016 Shaanxi Regional Statistical Yearbook (Statistics Bureau of Shaanxi Province, December 2016). The county-level administrative map was constructed using the National Earth System Science Data Sharing Infrastructure (http://www.geodata.cn). Natural geographic data were extracted, calculated, and analyzed from the 1 km Resolution Digital Elevation Model Dataset of China, obtained from the Scientific Data Center of Cold and Arid Regions (http://westdc.westgis.ac.cn). The population spatial database was established using ArcGIS 10.0.

The dependent variable in our models is the Regional Proportion of Population (RPP), calculated by dividing a county' s population by the total regional population. RPP effectively avoids heteroscedasticity caused by large area differentials between Shaanxi counties. Independent variables (economic and public service, population base, industrial structure, per capita income, topography, and average elevation) were obtained through factor analysis to avoid overlooking variables and multicollinearity issues.

The OLS regression model demonstrates a significant relationship between RPP and the explanatory variables, as well as dummy variables defined in the model. However, the dependent variable and its residuals cannot satisfy the no-spatial-autocorrelation assumption, although they satisfy other Gauss-Markov assumptions. Therefore, we also fit the spatial lag model (SLM) and spatial error model (SEM). Based on Lagrange multiplier (LM), Robust-LM, and Akaike information criterion tests, the SEM is the best model.

The SEM reveals that a county' s RPP in Shaanxi Province depends not only on observable variable characteristics but also on other characteristics that may be unobservable or omitted from the model. Among the independent variables, the economic and public service factor exhibits the most significant positive explanatory power on RPP. The population base factor, representing basic agricultural conditions, and the year 2000 population demonstrate positive explanatory power. The industrial structure factor is negatively correlated with RPP, indicating that the secondary industry has limited absorptive capacity for increasing employment compared with the tertiary industry. Average elevation has negative explanatory influence on RPP. The effects of per capita income and topography are insignificant, possibly because some counties with higher per capita income have economies based on natural resources or minerals and are often located in remote mountain areas. Nonetheless, topography exhibits a complex spatial coupling relationship with climate, precipitation, temperature,

and humidity. A county' s administrative rank significantly influences population distribution.

Our main conclusion is that the key, controllable determinative factors for optimizing population distribution are socioeconomic factors, although the restrictive role of natural geographic factors should not be overlooked. By considering spatial interactions between explanatory variables and error terms, this study corrects biased estimations of the OLS model and provides scientific analysis of factors influencing population distribution, which is of great reference value for projecting and optimizing population distribution trends.

Keywords: population distribution; influencing factors; regional proportion of population; spatial econometric model; Shaanxi Province

1. Data and Methodology

1.1 Data Sources and Variables

Geographic data were obtained from the National Earth System Science Data Sharing Infrastructure (http://www.geodata.cn). Natural geographic data were extracted from the 1 km resolution Digital Elevation Model dataset from the Scientific Data Center of Cold and Arid Regions (http://westdc.westgis.ac.cn). The explanatory variables include economic and public service factors, population base, industrial structure, per capita income, topography, and average elevation.

Descriptive statistics of original variables are presented in . The table shows the mean, standard deviation, minimum, and maximum values for each factor. The population base factor represents basic agricultural conditions, while the industrial structure factor reflects the proportion of secondary industry. Topographic factors include elevation and slope variables.

1.2 Spatial Econometric Models

The OLS model results indicate significant relationships but fail to account for spatial autocorrelation. Moran' s I test yields a value of 0.819 ($p < 0.001$), rejecting the null hypothesis of no spatial autocorrelation. The Jarque-Bera test statistic is 2.37 ($p > 0.05$), indicating residuals satisfy normality assumptions. However, spatial dependence tests (LM-lag = 11.929, $p < 0.001$; LM-error = 1.736, $p > 0.05$) suggest the spatial lag model is more appropriate.

We estimate both the spatial lag model (SLM) and spatial error model (SEM) as specified in equations (1) and (2):

Spatial Lag Model (SLM):

$$Y = \rho WY + X\beta + \varepsilon$$

$$\varepsilon \sim N(0, \delta^2 I_n)$$

Spatial Error Model (SEM):

$$Y = X\beta + \mu$$

$$\mu = \lambda W\mu + \varepsilon$$

$$\varepsilon \sim N(0, \delta^2 I_n)$$

where Y is the vector of RPP values, X is the $n \times k$ matrix of explanatory variables, W is the $n \times n$ spatial weight matrix, ρ is the spatial lag coefficient, β is the vector of coefficients, λ is the spatial error coefficient, and ε is the error term.

Model comparison results show the SEM provides the best fit with $R^2 = 0.9803$, Log-Likelihood = -134.63, and AIC = 289.27. The spatial error coefficient λ is statistically significant at the 1% level, confirming spatial dependence in the error term.

[Figure 1: see original paper] illustrates the regional proportion of population across Shaanxi counties in 2015, showing clear spatial clustering patterns. The spatial weight matrix W was constructed using queen contiguity, and all models were estimated using maximum likelihood estimation.

2. Results and Discussion

The SEM estimation results reveal several key findings. First, the economic and public service factor demonstrates the strongest positive effect on population distribution (coefficient = 0.671, $p < 0.001$). Counties with better public services and economic development attract larger populations. Second, the population base factor shows a positive coefficient (0.283, $p < 0.001$), indicating path dependency in population distribution.

Third, the industrial structure factor exhibits a negative coefficient (-0.660, $p < 0.001$), suggesting that counties dominated by secondary industry have lower population proportions, likely due to capital-intensive production requiring fewer workers. Fourth, average elevation has a negative effect (-0.42, $p < 0.001$), confirming that topographic constraints limit population settlement.

The administrative rank dummy variables are significant: D_1 (county-level city status) has a coefficient of 1.12 ($p < 0.001$), while D_2 (district status) shows -0.65 ($p < 0.001$). This indicates that administrative hierarchy significantly influences population distribution patterns.

The spatial error coefficient λ is 0.139 ($p = 0.007$), suggesting that unobserved factors create spatial clustering. The robust LM test confirms the SEM specification is appropriate (Robust LM-error = 6.932, $p = 0.031$).

3. Conclusion

This study demonstrates that spatial econometric models provide more reliable estimates than OLS for analyzing population distribution factors. Socioeconomic factors, particularly economic development and public services, are primary drivers of population distribution in Shaanxi Province. Natural geographic factors, especially elevation, impose significant constraints. Administrative rank plays an important role in shaping population patterns. These findings have important implications for regional planning and population policy optimization.

References

[1] Meng Xiangjing, Jia Shaofeng. The quantitative analysis of factors influencing population distribution in China. *Geographical Research*, 1993, 12(3): 56-63.

[2] Wang Li, Liu Ya, Cao Wentao, et al. Research on influential factors of population distribution in Hebei Province based on GWR model. *Journal of Liaoning Normal University (Natural Science Edition)*, 2016, 39(2): 262-267.

[3] Tong Yufen, Ma Yanlin. Influencing factors for the spatial distribution pattern of urban population: Take Beijing as an example. *Beijing Social Sciences*, 2016, (1): 89-97.

[4] Lv Chen, Fan Jie, Sun Wei. Population distribution and the influencing factors based on ESDA. *Economic Geography*, 2009, 29(11): 1797-1802.

[5] Gou Jinming, Yuan Tianfeng, Zhu Xiaoxia, et al. On the population distribution and influencing factors in northeast Sichuan. *Journal of Leshan Normal University*, 2013, 28(8): 100-103.

[6] Shen Silian, Wang Chunwei, Tang Jing. Research on the influencing factors of population distribution in Henan Province based on GWR model. *Mathematics in Practice and Theory*, 2014, 44(3): 165-174.

[7] Shi Ying, Mi Ruihua. Differentiation of population spatial distribution in Shaanxi Province. *Arid Land Geography*, 2015, 38(2): 368-376.

[8] Wang Siyan, Yang Chuanguo, Pang Hua, et al. Characteristics of population distribution in the Pearl River Basin and the impact factors analysis. *China Population, Resources and Environment*, 2014, 24(5): 447-450.

[9] Zhang Yaojun, Ren Zhengwei. Factors affecting population distribution in mountainous areas: Geographically weighted regression using data from Bijie. *Population Research*, 2012, 36(4): 53-63.

[10] Wang Gui, Feng Zhiming, Yang Yanzhao, et al. The change of population density and its influencing factors from 2000 to 2010 in China on county scale. *Acta Geographica Sinica*, 2014, 69(12): 1790-1798.

[11] Zhang Yaojun, Cen Qiao. Spatial pattern of population mobility and determinants of inter-provincial migration in China. *Population Research*, 2014, 38(5): 54-71.

[12] Liu Naiquan, Geng Wencai. On evolution of population spatial distribution pattern in Shanghai and its influencing factors: Empirical research based on spatial panel model. *Arid Land Geography*, 2015, 38(2): 368-376.

[13] Mu Xueying, Liu Kai, Ren Jianlan. Spatial pattern of the comprehensive carrying capacity of cities at prefectural level and above in China. *Arid Land Geography*, 2017, 40(3): 671-679.

[14] Wang Shaoping, Yang Jisheng, Ouyang Zhigang. Econometrics. Beijing: Higher Education Press, 2011: 61-113.

[15] Li Zhengsheng. Population Economics. 2nd ed. Beijing: Tsinghua University Press, 2009: 156-180.

[16] BURRIDGE P. On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1980, 42: 107-108.

[17] ANSELIN L. Spatial econometrics: Methods and models. Boston: Kluwer Academic Publishers, 1988.

[18] BERA AK, MCALEEER M, PESARAN MH, et al. Joint tests of non-nested models and general error specifications. *Econometric Reviews*, 1992, 11: 97-117.

[19] ANSELIN L. GIS research infrastructure for spatial analysis of real estate markets. *Journal of Housing Research*, 1998, 9: 113-133.

Note: Figure translations are in progress. See original paper for figures.