Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial dependence mediates polycentric effects on jobs–housing matching in three Chinese megacities

Juan Zhu
Chien-shiung Institute of Technology
Shengxin Chen
Tongji University
Wenqi Wang
Xi'an University of Architecture and Technology
Yao Wang
2689@mail.usts.edu.cn

Suzhou University of Science and Technology

Article

Keywords: jobs-housing matching, spatial econometric model, polycentric city, influencing factors, mobile signaling data

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11324484/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spatial dependence mediates polycentric effects on jobs–housing matching in three Chinese megacities

Juan Zhu [1], Shengxin Chen [2], Wenqi Wang [3] &Yao Wang [4]✉

How polycentric urban structures shape jobs–housing matching across cities at different developmental stages remains poorly understood. Using mobile signaling data from over 4.2 million commuters across Hangzhou, Wuhan, and Nanning—three Chinese megacities representing distinct phases of polycentric transition—this study constructs a grid-level jobs–housing matching rate and compares ordinary least squares with spatial econometric models (SEM/SLM). Results confirm that OLS residuals exhibit significant spatial autocorrelation, and spatial models substantially improve explanatory power. Critically, spatial dependence mechanisms differ across cities: Hangzhou and Wuhan follow the spatial error model, while Nanning follows the spatial lag model. Distance to the nearest employment center consistently and negatively affects matching in larger cities, but positively in Nanning's periphery. Land use mix benefits worker matching but negatively affects resident matching in Hangzhou and Wuhan, suggesting functional mixing may crowd out residential space. Metro accessibility shows limited or negative effects on resident matching across all three cities, likely due to transit capitalization raising local housing costs. The study identifies three distinct urban regimes—polycentric polarization with housing-price crowding-out (Hangzhou), uneven polycentricity with strong spatial dependence (Wuhan), and peripheral cluster self-sufficiency with spatial spillovers (Nanning)—highlighting that achieving jobs–housing balance requires spatially aware, context-specific planning strategies tailored to each city's developmental stage.

Keywords jobs-housing matching, spatial econometric model, polycentric city, influencing factors, mobile signaling data

Whether polycentric urban structures improve jobs–housing balance and shorten commutes remains a central debate in urban geography and transportation research. The "co-location hypothesis," proposed by Gordon et al. [1], argues that both households and firms adjust their locations to avoid congestion, thereby reducing commuting distances. Empirical evidence from the United States [1–3], the Netherlands [4], Turkey [5], and China [6–9]. However, critics raise two objections: first, residential rigidity makes commuting the default response [10]. Second, asynchronous suburbanization of population and employment creates low-self-containment centers that may increase commuting costs [11–14]. Even with polycentric employment, strong CBD agglomeration may perpetuate centripetal commuting and mismatch [15].

Empirical evidence remains polarized. Inter-city comparisons yield conflicting results: some studies report longer commute times in high-density monocentric U.S. cities [1], with subsequent analyses providing corroborating evidence [8,16]. Others, however, reach opposite conclusions [14,17,18]. Intra-city studies are equally divided. Supportive studies [2–7,19,20] find that employment polycentricity shortened commute times in Los Angeles [2], improves job accessibility in the Dutch Randstad [4], and promotes sustainable transport in Hong Kong [6]. In contrast, opposing studies [11–13,21–23] document rapid vehicle-miles traveled growth and low self-containment in suburban centers [21–23]; longer commutes in both traditional and emerging centers in Beijing [12]; and a 10.4% reduction in commute times for workers in inner-city subcenters versus a 27.4% increase for those in suburban subcenters in Seoul [13].

Beyond the mere presence of polycentricity, the degree of polycentricity also exerts significant influence [10,16,24–26]. For instance, research across 82 Italian metropolitan areas finds that higher polycentricity correlated with greater private transport use and higher external commuting costs [10]. A study of 50 U.S. metropolitan areas demonstrates that employment density around strongly polycentric centers reduced commute time, but increases it around moderately polycentric centers [26]. Similarly, larger employment centers are associated with longer commute times in Atlanta [24]. Conversely, a study of seven Spanish urban systems reports that polycentricity is negatively correlated with commuting distance, and that a more equitable distribution of employment across hierarchical levels further shortens commutes [16]. In Dutch metropolitan regions, different polycentric configurations exert inconsistent effects on jobs–housing spatial relationships [25]. Collectively, these findings suggest that the relationship between polycentricity and jobs–housing balance is contingent on the specific configuration and maturity of polycentric development.

[1] Suzhou Chien-shiung Institute of Technology, College of Art and Design, Suzhou, China.[2]Tongji Architectural Design (Group) Co., Ltd., Tongji University, Shanghai, China.[3] Design and Research Institute Co., Ltd., Xi'an University of Architecture and Technology, Xi'an, China.[4]College of Architecture and Urban Planning, Suzhou University of Science and Technology, Suzhou, China. ✉email: 2689@mail.usts.edu.cn

Apart from urban spatial structure, macro-structural, meso-level built environment, and micro-level individual factors collectively shape jobs–housing relationships [2,27–29]. Macro-structural factors include economic structure, as different economic sectors generate distinct commuting patterns [30–32]; housing costs ,which constrain jobs–housing matching through market dynamics [21,29,33–35], and institutional contexts, exemplified by China's danwei system and the post-1998 housing reform, which created a growing disconnect between employment centers and residential populations [36]. At the meso-level, built environment attributes such as land use mix [37–39], population and employment density [5,13,14], job accessibility [16,24,37], public service facility [16,39], and relevant planning policies [7,19] moderate commuting behavior beyond the effects of polycentricity. At the micro level, individual and household socioeconomic attributes—including income, occupation, family structure, educational attainment, and gender—significantly influence jobs–housing spatial relationships. Residential location decisions are typically determined by multiple factors beyond workplace proximity, such as neighborhood amenities, safety, and income levels [2,27]. Indeed, micro-level attributes often exert stronger influences on commuting outcomes than urban form itself [2,14,17,27,31].

China's urban spatial restructuring differs fundamentally from Western experiences. Unlike the "generalized decentralization" model prevalent in North America, China's polycentric development approximates a pattern of "decentralized concentration"—that is, emerging employment subcenters have not fully supplanted the dominance of traditional central urban areas [40–42]. Manufacturing suburbanization, rather than tertiary-sector dispersal, has been the primary driver of this process, alongside strong government planning interventions [43,44].Moreover, China's unique institutional background—including the hukou (household registration) system, danwei (work unit) system, housing marketization reforms, and industrial policies—has continuously reshaped jobs-housing spatial relationships [7,12,19,31]. Under these distinctive conditions, the mechanisms, magnitude, and even direction of polycentric structures' effects on jobs-housing balance may exhibit complex variations that remain insufficiently understood. Systematic comparative research across different types and stages of Chinese cities within a unified analytical framework is notably lacking.

To address this gap, this study analyses how polycentric spatial structure and built environment factors shape jobs–housing matching for both workers and residents in three Chinese megacities at distinct developmental stages: Hangzhou, Wuhan, and Nanning. The analysis focuses on three core dimensions of the urban environment—namely, polycentric spatial structure, land use composition, and transportation accessibility—measured at the grid-cell level to capture the spatial heterogeneity of jobs–housing relationships. Methodologically, the study integrates Ordinary Least Squares (OLS) regression to establish baseline estimates, spatial econometric models (Spatial Error Model and Spatial Lag Model) to account for spatial dependence and spillover effects, and a systematic comparison across model specifications to identify the most appropriate spatial process. This study contributes to a deeper understanding of how urban form interacts with commuting efficiency, reinforcing the importance of context-sensitive, spatially-aware planning interventions as levers for enhancing jobs–housing balance, transportation sustainability, and urban equity across diverse developmental contexts.

Materials and methods

Study area

To capture regional development differences across distinct stages of polycentric transition, three Chinese megacities were selected based on population size, economic development level, and spatial structure characteristics: Hangzhou (8.57 million; per capita GDP ¥135,877), Wuhan (11.081 million; ¥187,263), and Nanning (4.418 million; ¥ 57,781) (National Bureau of Statistics, 2021). These cities represent a gradient of polycentric maturity—from the well-established polycentric structure of Hangzhou in the Yangtze River Delta, to the transitional phase of Wuhan in Central China, to the nascent polycentric development of Nanning in the western region.

Given the considerable variation in administrative areas (Hangzhou 16,850 km², Wuhan 8,494 km², Nanning 22,099 km²), scale definition is critical for meaningful polycentricity measurement. Following Parr [45], the study adopted Functional Urban Regions (FURs) by aggregating subdistricts meeting the following criteria: (1) ≥5% commuting flow to the urban core, (2) above-average population density, (3) population >5,000, and (4) road-network clustering area >5 km². The resulting FURs yielded comparable study areas of 4,091.88 km² (Hangzhou, 24.28% of administrative area), 5,026.96 km² (Wuhan, 59.18%), and 4,827.73 km² (Nanning, 21.85%), ensuring spatial comparability across the three cities (Figure 1).

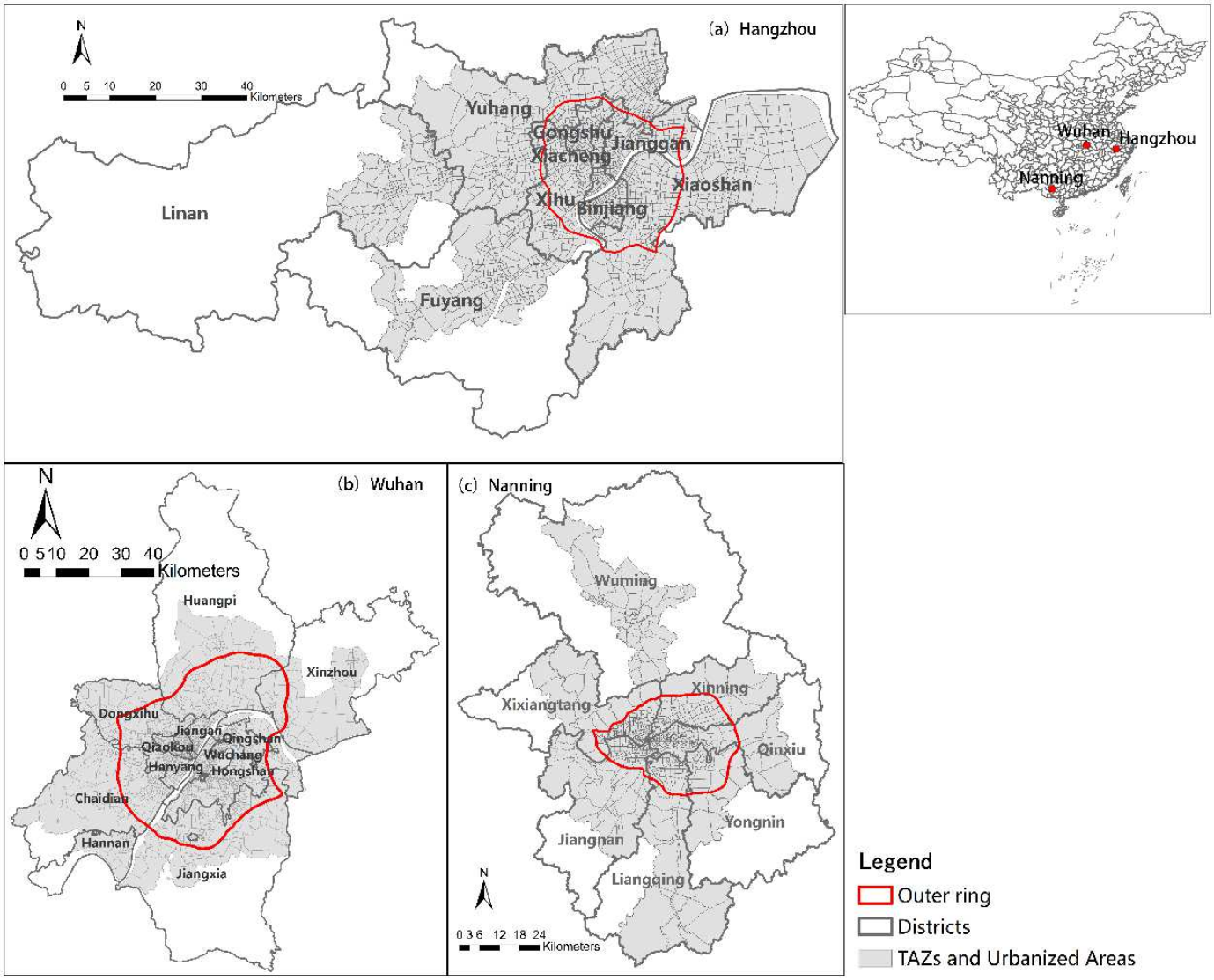


Fig. 1. Study Area (The inset map in the upper right corner shows the geographical locations of the three cities within China, and the left panel shows the study areas of the three cities:(a) Hangzhou, (b) Wuhan, and (c) Nanning) Mapping based on software ArcGIS pro 3.3.

Study data

Mobile signaling data

This study measures jobs–housing spatial matching using commuting populations identified from mobile signaling data. This study used anonymized and aggregated mobile signaling data provided by China Mobile. The data were pre-processed by the provider to remove all personally identifiable information and were aggregated at the 1 km grid level prior to analysis. This study uses continuous one-month China Mobile signaling data from March, April, and August 2017 for Nanning, Hangzhou, and Wuhan, respectively; using data from the same year ensures cross-city comparability. Following the identification method of Zhu et al. [46,47], a total of 2.2283 million, 1.1986 million, and 824,900 commuting individuals were identified within the administrative areas of Hangzhou, Wuhan, and Nanning, respectively, of which the study areas accounted for 88.47%, 98.57%, and 69.28% of each city's total commuting population.

OpenStreetMap (OSM) data

OpenStreetMap (OSM) data refers to the collaborative, freely accessible geospatial database of global transportation infrastructure contributed by volunteers and authoritative sources. By extracting road networks, subway lines, and major transport nodes, OSM data can characterize urban transport supply and accessibility patterns. This study uses OSM data for 2017, including metro stations, outer ring roads, major infrastructure (nearest airports, railway stations, and highway interchanges), and major natural barriers (major rivers, lakes, or scenic areas)—to assess transportation accessibility.

Amap(Gaode) POI data

Amap Points of Interest (POI) data refers to georeferenced point features representing various public service facilities, collected through the *Gaode* mapping platform. By analyzing the spatial distribution and density of, such data can characterize the provision and spatial pattern of urban public services. This study uses Amap POI data for 2018, including four major categories of public service facilities: government agencies and social organizations, healthcare services, sports and leisure services, and science, technology, and cultural services.

A total of 86,163, 72,102, and 8,751 records were collected within the municipal areas of Hangzhou, Wuhan, Nanning, respectively.

Land use data

Land use data refers to the spatial classification of surface cover types, obtained from government planning departments and remote sensing-based classification datasets. By distinguishing between different land use categories—such as residential, commercial, industrial, and green space—land use data can characterize the balance of land use mix. This study uses current land use data for Hangzhou and Nanning obtained from government planning departments, and adopts the national land use classification dataset developed by Gong[48] et al. for Wuhan. To integrate these two sources, the urban construction land data were merged with Gong et al.'s dataset, ultimately forming eight major land use types: residential, commercial, business office, industrial, transportation, public service facilities, educational and scientific research, and parks and green land. Through preprocessing, the land use polygons were rasterized and aggregated to the 1 km grid cells.

Housing price data

Housing price data refers to the transaction or listing prices of residential properties, obtained through web crawling technology from online real estate platforms. This study uses housing price data for 2020 obtained through web crawling from SouFun.com. A total of 4,750, 4,430, and 4,182 records were collected for Hangzhou, Wuhan, and Nanning, respectively. Each record includes community name, unit price, building area, land area, total number of households, greening rate, plot ratio, and WGS84 longitude and latitude coordinates. Through preprocessing, duplicate records were removed, outliers were trimmed, and the data were spatially joined to the 1 km grid cells.

Urban construction land time-series data

Urban construction land temporal sequence data refers to the annual spatial extent of built-up land derived from long-term remote sensing observations. By tracing the expansion of construction land over decades, such data can reveal the historical process of urban sprawl and land development intensity. This study adopts the national urban construction land temporal sequence dataset (1978–2017) developed by Gong et al. [49]. The construction start year was set as the year prior to 1985 (i.e., the construction age for land developed before 1985 was uniformly set to 33 years, and the construction age for 2017 was set to 1 year). Through preprocessing, the vector polygons were rasterized and assigned to the 1 km grid cells.

Methods

Identifying employment centers

This study compares three methods for identifying employment centers: the absolute threshold method of Giuliano and Small's [50], the relative threshold method of Muniz et al.'s [51], and the jobs-housing ratio method of Shearmur and Coffey's [52], each with distinct strengths and limitations. Adopting the approach of Zhu et al. [46], this study employs a combined method that integrates the relative threshold approach with the jobs–housing ratio. Within the central urban area (inside the outer ring road), employment centers are identified using Equations (1), (2), and (3) jointly:

$$D_{taz} \geq \overline{D} \quad (1)$$

$$E_{taz}/R_{taz} \geq 1; \quad (2)$$

$$E_t \geq 0.5\%\mathrm{E} \quad (3)$$

Outside the central urban area, employment centers are identified using either Equations (1) and (5), or Equations (4), (2), and (5) (identification results are shown in Figures 2):

$$D_{taz} \geq \overline{D_{out}} \quad (4)$$

$$E_t \geq 0.25\%\mathrm{E} \quad (5)$$

where D_{taz} is the employment density of the TAZ unit (person /km^2), $\overline{D}$ is the average employment density of all the TAZ units, E_{taz} is the employment population of the TAZ unit, R_{taz} is the resident population of the TAZ unit, E_t is the total employed population of the adjacent units (queens rule is adopted), E is the total employed population of all the TAZ units, and $\overline{D_{out}}$ is the average employment density of all the TAZ units outside the central urban area.

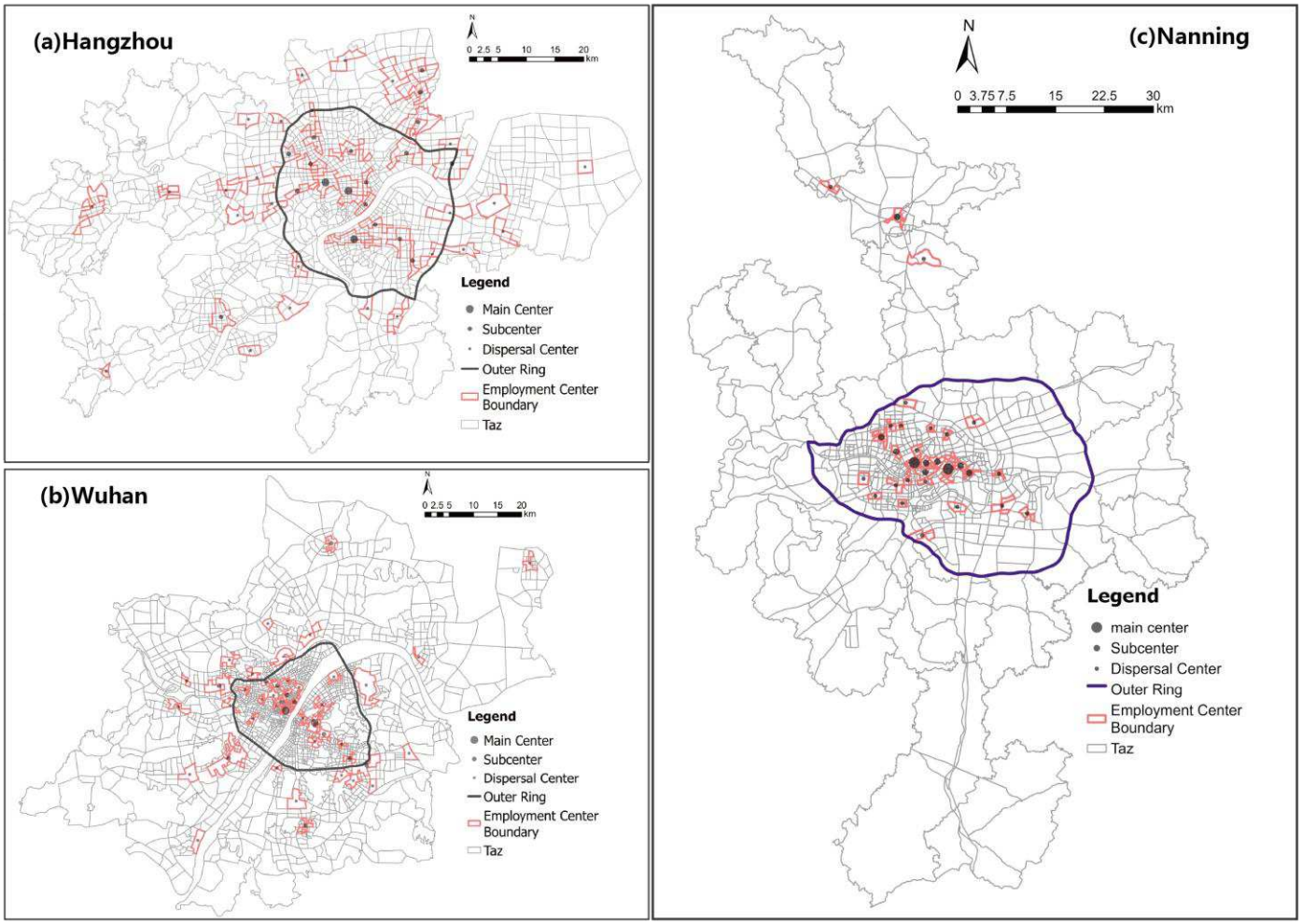


Fig. 2. Identification of employment centers (The figures show the employment centers of the three cities:(a) Hangzhou, (b) Wuhan, and (c) Nanning) Mapping based on software ArcGIS pro 3.3.

Measuring jobs–housing *matching*

Existing studies have quantified jobs–housing balance using ratio-type indicators (e.g., jobs-to-housing-unit ratio, JHR [23,53]; jobs-to-household ratio, JHR [28,33]; jobs-to-population ratio, JPR [28]; jobs-to-employed-resident ratio, JER, or jobs-to-workers-residing ratio, JWR [38,54];independence index [21,38]) and commuting behavior indicators (e.g., commuting time [1,3,5,8,12,13,15,17,26]; commuting distance [11,25,55]). However, these indicators face methodological limitations centered on three issues: sensitivity to spatial unit scale [56], arbitrary boundary delineation [54], and questionable validity in representing true proximity.

To address these challenges, this study integrates Cervero's [33] sub-region matching concept (e.g., a 3–5 mile radius from the workplace) with Peng's [28] actual commuting distances approach, this study constructs a "job–housing matching rate": within a spatial unit, the proportion of commuters whose commuting distance falls within a reasonable threshold (here, the city-wide average commuting distance from mobile signaling data) This indicator can be disaggregated into worker, residential, and overall matching rates. The specific calculation method is illustrated in Figure 2, with the formulas as follows:

$$\text{Workers' jobs-housing matching rate} = (A+B')/(A+B'+C')$$

$$\text{Residential jobs-housing matching rate} = (A+B)/(A+B+C)$$

$$\text{Overall jobs-housing matching rate} = (A+B+B')/(A+B+C+B'+C')$$

This study uses 1 km grid cells rather than TAZs to measure jobs–housing matching. Uniform grids mitigate the MAUP caused by varying TAZ sizes, ensuring stable and comparable results. This scale matches mobile signaling data precision (average base station densities: 9.52, 1.38, 0.65 stations/km² in Hangzhou, Wuhan, Nanning), avoiding data voids at finer scales (e.g., 500 m). Additionally, 1 km grids reveal fine-grained spatial differences, preventing the masking of local features by coarser units (e.g., TAZs > 2 km). The calculation results are shown in Figure 3.

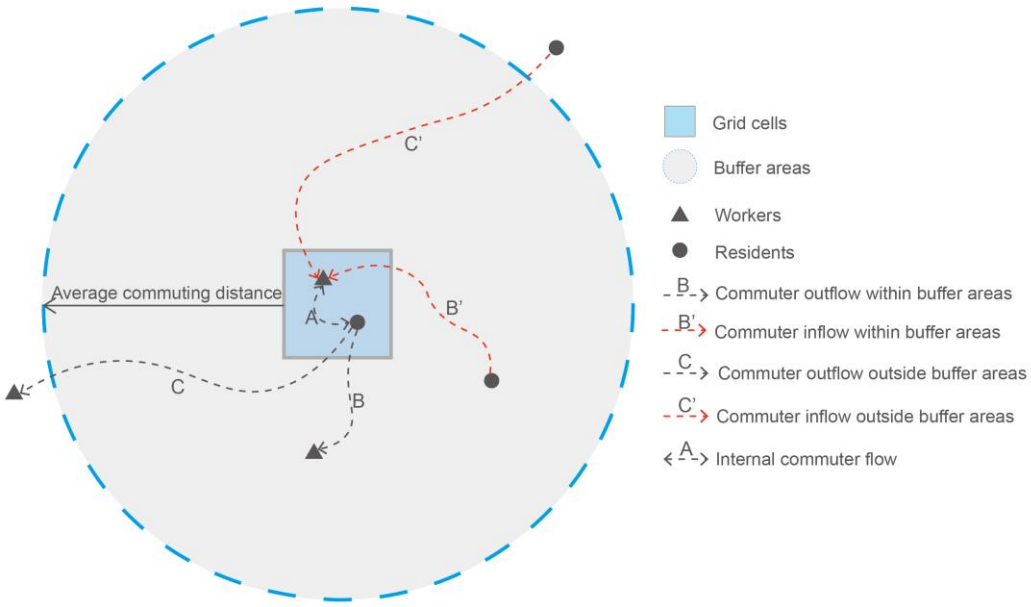


Fig. 2. The jobs-housing matching index calculation rules based on commuting flow



Fig 3. The grid-level jobs-housing matching rates (Each row displays the study areas across three dimensions: (a) Workers, (b) Residents, and (c) Overall, illustrating the spatial variation of job-housing matching within three cities:(a) Hangzhou, (b) Wuhan, and (c) Nanning) Mapping based on software ArcGIS pro 3.3.

Analysis of factors influencing jobs–housing matching

This study adopts a three-step framework—theoretical construction, model selection, and factor decomposition—to examine the determinants of jobs–housing spatial matching. First, an ordinary least squares (OLS) baseline model is established. Spatial regression models, namely the spatial lag model (SLM) and the spatial error model (SEM), are specified by introducing a spatial lag term or a spatial error term, respectively, to address the bias of OLS estimates under spatial autocorrelation. Model selection is then guided by a series of spatial diagnostics, including tests for spatial dependence and model comparison criteria.

The dependent variable is the jobs–housing matching rate for residents and workers in grid cell i, denoted $Blance_i$. The independent variables are organized into two dimensions: (1) macro-structure, i.e., urban spatial structure factors, denoted $Spatial_i$; and (2) micro-environment, i.e., local environmental factors, denoted $Local_i$. The latter encompasses the neighborhood-built environment ($Neighb_i$), grid-level accessibility ($Access_i$), and urban economic development factors (Z_i) (Table 1-2). The final equation is as follows:

$$Balance_i = f(Spatial_i, Local_i) + u_i = f(Spatial_i, Neighb_i, Access_i, Z_i) + \varepsilon_i$$ (230)

where ε_i is the error term. (231)

Urban spatial structure ($Spatial_i$) examines the impact of overall urban form on jobs–housing relationships through three variable sets: polycentric spatial structure, industrial structure, and polycentricity level. Polycentric spatial structure is measured by the distances to the nearest employment center ($D_{cen,i}$), primary center ($D_{main,i}$), subcenter ($D_{sub,i}$), and dispersed center ($D_{fensan,i}$), testing decay effects of agglomeration economies at different scales[13,24]. All distances are calculated as Euclidean distances between grid centroids and the corresponding centers, using proximity analysis in ArcGIS software. Industrial structure is captured by the Hirschman–Herfindahl Index (HHI_i) and relative specialization index (Spe_i), calculated based on economic census or land use data. Both indicators are distance-weighted to the nearest center to incorporate a distance decay effect. Higher agglomeration implies stronger employment attraction, while higher specialization indicates more homogeneous employment opportunities—both of which tend to worsen jobs–housing matching [31]. Polycentricity level is measured by employment population size (Emp_i), employment density (Den_i), and functional polycentricity indices for workers and residents (E_{Rsqu_i}、H_{Rsqu_i}) [57], also distance-weighted. Areas with more jobs and higher density tend to produce longer commutes [13,14].

Neighborhood built environment ($Neighb_i$) captures micro-level facility and service provision based on POI data, covering land use and public services. Land use is measured by mix entropy (ENT_i) and residential-non-residential balance (BAL_i) within the grid's average commuting buffer [58]; higher mix reduces commuting distance [38]. Public services are measured by POI densities of government (Pub_i), education (Edu_i), sports ($Sport_i$), and healthcare ($Hosp_i$) within the same buffer; better services facilitate matching [16].

Accessibility ($Access_i$) includes distances to the nearest metro station ($D_{metro,i}$), outer ring road ($D_{huanxian,i}$), and major infrastructure ($D_{infra,\ i}$) — airport, railway station, highway interchange. Better accessibility facilitates matching [24,37]. Natural barriers ($D_{hinder,i}$) measured as distance to the nearest river, lake, or scenic area, are hypothesized to hinder commuting.

Urban economic development (Z_i) includes average housing price ($Price_i$) and average building age ($Year_i$). High prices and suburban housing shortages worsen mismatch [21,35]; affordable housing near centers aids low- and middle-income or single workers [29]. Longer construction history implies more stable populations and better jobs–housing matching, while newer centers undergoing functional reorganization tend to exhibit poor matching [12]. Construction duration is captured by the average building age of construction land ($Year_i$), calculated from Gong et al.'s [49] 1978–2017 land use dataset.

Variables	Description
Dependent Variables	
Worker $Balance_i$	The jobs-housing matching rate of workers.
Resident $Balance_i$	The jobs-housing matching rate of residents.
Independent Variables	
Urban Spatial Structure	
$D_{cbd,i}$	Distance from grid i to the nearest main center(km)
$D_{sub,i}$	Distance from grid i to the nearest sub-center(km)
$D_{fensan,i}$	Distance from grid i to the nearest dispersed employment center(km)
$D_{cen,i}$	Distance from grid i to the nearest employment center(km)
$HHI_{cen,i}$	$HHI_{cen,i}=\frac{\sum(q_{cen,j}/q_{cen})^2}{D_{cen,i}}$, $q_{cen,j}$ represents the industrial employment population in industry j of the nearest employment center to grid i; q_{cen} represents total employment populations across all industries of the nearest employment center to grid i. For Wuhan and Nanning, where economic census data are unavailable, land use data are used as a substitute.
$Spe_{cen,i}$	$Spe_{cen,i}$=max($LQ_{cen,j}$)/ $D_{cen,i}$ = max($\frac{\frac{q_{cen,j}}{q_j}}{\frac{q_{cen}}{q}}$) / $D_{cen,i}$, $LQ_{cen,j}$ represents the industrial location quotient of the nearest employment center to grid i; q_j represents the total employment population in industry j across all centers; q represents the total employment population of all industries in the city. For Wuhan and Nanning, where economic census data are unavailable, land use data are used as a substitute.
$Emp_{cen,i}$	Employment population size of the nearest employment center for grid i/$D_{cen,i}$
$Den_{cen,i}$	Employment population density of the nearest employment center for grid i/$D_{cen,i}$
$E_{Rsqu_{cen,i}}$	Worker R^2 of the nearest employment center for grid i/$D_{cen,i}$
$H_{Rsqu_{cen,i}}$	Resident R^2 of the nearest employment center for grid i/$D_{cen,i}$
Neighborhood Built Environment	

$ENT_{i,j}$	$ENT_{i,j}$=$\frac{-\sum_{j=1}^{k} p_{i,j} \ln p_{i,j}}{\ln k}$, k denotes the number of land use types within the buffer zone of the average commuting distance of grid i. $p_{i,j}$ denotes the area proportion of the j-th land use type within the buffer zone of the average commuting distance of grid i.
BAL_i	BAL_i=1- $\frac{\lvert X_i - aY_i \rvert}{\lvert X_i + aY_i \rvert}$, X_i and Y_i represent the proportions of residential and non-residential land, respectively, within the buffer zone of the average commuting distance of grid i. a is the balance adjustment factor, which is set to 1 in this study, based on the assumption that within the metropolitan area, residential and non-residential land use is essentially balanced across Hangzhou, Wuhan, and Nanning.
Pub_i	Number of government agencies and social organizations POIs in the average commuting ditance buffer zone of grid i/Buffer zone area(counts/km²)
Edu_i	Number of science, education, and cultural services POIs in the average commuting ditance buffer zone of grid i/Buffer zone area(counts/km²)
$Sport_i$	Number of sports and leisure services POIs in the average commuting ditance buffer zone of grid i/Buffer zone area(counts/km²)
$Hosp_i$	Number of healthcare services POIs in the average commuting ditance buffer zone of grid i/Buffer zone area(counts/km²)
Accessibility	
$D_{metro,i}$	Distance from grid i to the nearest metro station(km)
$D_{huanxian,i}$	Distance from grid i to the outer ring road(km)
$D_{infra,\ i}$	Distance from grid i to the nearest major infrastructure(km)
$D_{hinder,i}$	Distance from grid i to the nearest large natural barrier(km)
Urban Economic Development	
$Price_i$	Total housing price of all houses within the average commuting radius of grid i/Total number of houses
$Year_i$	(Number of cells in grid I *cell age)/Total number of cells

Table 1. Definition and Classification of Variables

Variables	Hangzhou Descriptive Statistics (*N* =2660)				Wuhan Descriptive Statistics (*N* = 1830				Nanning Descriptive Statistics (*N* = 1050)			
	Mean	S. D.	Min	Max	Mean	S. D.	Min	Max	Mean	S. D.	Min	Max
Dependent Variables												
Worker $Balance_i$	0.74	0.16	0	1	0.70	0.17	0	1	0.67	0.19	0	1
Resident $Balance_i$	0.73	0.16	0	1	0.71	0.16	0	1	0.69	0.18	0	1
Independent Variables												
Urban Spatial Structure												
$D_{cbd,i}$	20.95	12.31	0	56.95	19.70	12.07	0	58.19	20.44	15.78	0	69.33
$D_{sub,i}$	11.07	8.74	0	40.47	10.38	8.75	0	44.88	14.82	13.47	0	62.69
$D_{fensan,i}$	5.86	3.90	0	20.10	58.37	5.82	0	33.53	10.37	11.50	0	54.58
$D_{cen,i}$	4.26	3.60	0	20.10	4.68	4.65	0	33.53	11.67	4.94	0	54.58
$HHI_{cen,i}$	0.18	0.30	0.01	8.70	0.25	0.45	0.02	8.90	0.40	0.45	0.04	5.17
$Spe_{cen,i}$	1.03	1.88	0.11	48.02	4.37	12.77	0.06	367.94	2.48	4.05	0.14	40.87
$Emp_{cen,i}$	9043.60	25756.84	240.70	947436.44	4812.50	9133.05	147.92	158295	4984.24	3868.18	1497	24348
$Den_{cen,i}$	884.47	2057.43	35.77	61424.68	1146.97	2777.77	29.83	42087.24	2007.06	1313.97	218.26	8300.46
$E_{Rsqu_{cen,i}}$	0.02	0.10	0	3.82	0.02	0.06	0	1.27	0.09	0.12	0	0.54
$H_{Rsqu_{cen,i}}$	0.05	0.17	0	6.34	0.04	0.10	0	1.97	0.11	0.15	0	0.69
Neighborhood Built Environment												
$ENT_{i,j}$	0.53	0.16	0	0.79	0.56	0.17	0	0.85	0.59	0.15	0	0.84
BAL_i	0.59	0.29	0	1	0.53	0.33	0	1	079	0.21	0	1

Pub_i	5.56	9.82	0	70.20	7.20	10.47	0.04	53.88	3.41	4.80	0.01	18.05
Edu_i	6.34	11.46	0	75.21	9.68	14.35	0.01	64.33	4.32	5.04	0.01	19.66
$Sport_i$	3.92	5.63	0	34.75	5.66	8.27	0.01	37.96	3.95	5.09	0.01	18.62
$Hosp_i$	3.99	6.03	0	39.30	6.31	9.23	0.01	44.28	4.77	5.67	0.01	21.24
Accessibility												
$D_{metro,i}$	16.97	13.86	0.04	60.20	10.45	9.91	0.05	48.35	16.77	15.88	0	63.10
$D_{huanxian,i}$	11.80	10.29	0.01	47.66	11.04	9.92	0	47.12	12.38	12.88	0	59.73
$D_{infra,\ i}$	4.76	3.49	0.11	23.91	4.28	2.87	0.07	24.31	7.03	8.07	0	42.46
$D_{hinder,i}$	5.48	4.81	0	23.34	2.83	3.65	0	22.74	10.16	13.35	0	53.81
Urban Economic Development												
$Price_i$	23689.90	8594.90	8070	52221.18	14549.19	4720.91	2444	53503	9108.67	2957.90	3899	16358.8
$Year_i$	12.18	5.85	1	34	13.45	6.98	1.5	32.39	17.40	7.13	1	32

Table 2. Descriptive statistics of variables

Given the large number of candidate variables, partial correlations analysis was first conducted to detect multicollinearity. Variable combinations were then estimated progressively, retaining those with stable and statistically significant coefficients. Spatial econometric diagnostics were subsequently performed on the OLS baseline model using the Lagrange Multiplier (LM) test in GeoDa, including LM-Lag, Robust LM-Lag, LM-Error, Robust LM-Error, and LM-SARMA statistic. Model selection followed established rules: (1) if neither LM statistic is significant, the OLS model is retained; (2) if only one LM statistic is significant, the corresponding spatial model (SLM or SEM) is adopted; (3) if both LM statistics are significant, the model with the more significant Robust LM statistic is selected. The functional forms of the SEM and SLM are specified as follows:

$$Balance_i = \mathrm{f}(Polycen_i, Neighb_i, Access_i, Z_i) + \lambda W\varepsilon_i + u_i \quad (SEM)$$

$$Balance_i = \mathrm{f}(WBalance_i, Polycen_i, Neighb_i, Access_i, Z_i) + \varepsilon_i \quad (SLM)$$

Where W is the spatial adjacency weight matrix, λ is the spatial error correlation coefficient, $W\varepsilon_i$ represents the spatial error affected by neighboring regions, and $WBalance_i$ denotes the spatial lag of the dependent variable, i.e., the jobs–housing matching rate affected by neighboring regions.

Results

Multicollinearity and stability tests of the OLS model

To test model robustness, a combined stepwise regression and variable screening approach was employed, sequentially building three nested models: Model 1 (neighborhood environment and control variables), Model 2 (adding transportation accessibility), and Model 3 (adding polycentric variables). Multicollinearity was evaluated using partial correlations, variance inflation factors (VIF), and coefficient stability (Table 3).

In Hangzhou, high correlations among public service variables (e.g., $Sport_i$ with $Hosp_i$: 0.77; $Sport_i$ with Edu_i: 0.72; Pub_i with $Hosp_i$: 0.72; Pub_i with Edu_i:0.56) led to removal of $Sport_i$ and $Hosp_i$ in Model 1. In Model 2, Pub_i and Edu_i correlated at 0.88, prompting the removal of Pub_i. In Model 3, despite some elevated correlations (e.g., $E_{Rsqu_{cen,i}}$ with $H_{Rsqu_{cen,i}}$: 0.69; $D_{huanxian,i}$ with $D_{main,i}$: 0.69), most VIF values remained acceptable, and theoretically important variables were retained. Stepwise regression showed stable coefficients and significance, confirming model robustness.

In Wuhan, $Sport_i$ and $Hosp_i$ were removed due to high correlations (e.g., $Sport_i$ with $Hosp_i$: 0.65; $Sport_i$ with Edu_i: 0.90; $Pub_i$with $Hosp_i$: 0.63). Model 2 variables showed low correlations (all VIF < 10). In Model 3, $D_{main,i}$ and $D_{huanxian,i}$ correlated at 0.80 (VIF: 41 and 23), so $D_{main,i}$ was excluded. Remaining coefficients remained stable.

In Nanning, $Sport_i$, Pub_i, and $Hosp_i$ were removed due to substantial overlap (e.g., $Sport_i$ with Pub_i: 0.91; $Sport_i$ with Edu_i : 0.47; $Sport_i$ with $Hosp_i$: 0.41; Pub_i with $Hosp_i$: 0.43; $Hosp_i$ with Edu_i:0.51). After removal, model 2 VIF values were all below 10. In Model 3, $D_{sub,i}$ correlated with $D_{main,i}$ (0.59) and $D_{metro,i}$ (0.75), with a VIF of 23, leading to its exclusion. Remaining variables, despite some high correlations, had VIF values near 10 and were retained. Coefficients remained stable, indicating effective multicollinearity control.

Variable	Hangzhou			Wuhan			Nanning		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Constant	0.714*** (41.26)	0.818*** (42.30)	0.859*** (27.48)	0.637*** (28.83)	0.602*** (24.77)	0.470*** (10.982)	0.811*** (12.24)	0.693*** (12.62)	0.521*** (6.25)
$D_{main,i}$			8.912E-7			1.388E-5***			$2.518E-5$***

			（1.06）			（-6.075）			（4.12）
$D_{sub,i}$			$-1.414E-6^{*}$ （-1.84）			2.447E-6** （1.98）			$-1.593E-5^{**}$ （-2.06）
$D_{fensan,i}$			-6.798E-8 （-0.04）			2.576E-6 （1.14）			$-2.002E-5^{**}$ （-2.14）
$D_{cen,i}$			-0.007^{***} （-3.16）			-0.011^{***} （-4.00）			6.988E-7 （0.06）
$HHI_{cen,i}$			-0.082^{***} （-4.52）			0.026 （1.39）			-0.021 （-0.89）
$Spe_{cen,i}$			0.000 （-0.05）			0.000 （0.83）			-0.006^{**} （-2.42）
$Emp_{cen,i}$			3.441E-7* （1.66）			$-1.348E-6^{**}$ （-2.19）			-2.182E-6 （-0.95）
$Den_{cen,i}$			$1.236E-5^{***}$ （3.12）			8.577E-7 （0.27）			1.029E-5* （1.69）
$E_{Rsqu_{cen,i}}$			-0.203^{**} （-2.26）			-0.825^{***} （-5.60）			-0.083 （-0.86）
$H_{Rsqu_{cen,i}}$			0.006 （0.12）			0.260^{**} （2.49）			0.019 （0.62）
$ENT_{i,j}$	0.174^{***} （5.48）	0.234^{***} （7.68）	0.161^{***} （4.88）	-0.022 （0.68）	-0.011 （-0.33）	-0.055 （-1.58）	-0.214^{***} （2.73）	-0.087 （-1.17）	-0.153^{**} （-1.87）
BAL_i	-0.007 （-0.57）	0.051^{***} （3.55）	0.051^{***} （3.40）	0.080^{***} （4.81）	0.091^{***} （536）	0.125^{***} （6.51）	0.009 （0.20）	0.027 （0.61）	0.083 （1.48）
Pub_i	-0.008^{***} （-6.02）	0.000 （0.27）		-0.008^{***} （-3.56）	0.000 （-0.22）	0.003^{***} （3.47）	-0.020^{**} （-2.20）		
Edu_i	-0.004^{***} （-2.90）	-8.050E-5 （-0.10）	-0.001 （-1.29）	0.006^{***} （4.43）	0.001 （1.44）	0.003^{***} （4.09）	0.003 （0.33）	0.010^{***} （4.49）	0.013^{***} （4.13）
$Sport_i$	0.001 （0.33）			-0.020^{***} （-4.22）			0.013 （0.89）		
$Hosp_i$	0.018^{***} （5.92）			0.021^{***} （6.27）			0.008 （1.40）		
$D_{metro,i}$		0.000 （-0.28）	$-1.422E-5$ （-0.03）		-0.001 （-1.06）	-0.003^{***} （-2.69）		0.005^{*} （1.85）	0.001 （0.34）
$D_{huanxian,i}$		-0.003^{***} （-4.90）	-0.003^{***} （-2.69）		0.002^{**} （1.99）	-0.008^{***} （-3.81）		-0.009^{**} （-2.24）	-0.006 （-1.50）
$D_{infra,\ i}$		-0.005^{***} （-4.44）	-0.005^{***} （-4.55）		0.000 （0.23）	-0.001 （-0.69）		0.006^{*} （1.53）	0.005 （1.16）
$D_{hinder,i}$	0.005^{***} （6.33）	0.004^{***} （5.63）	0.006^{***} （7.39）	0.010^{***} （7.47）	0.009^{***} （5.25）	0.002 （0.96）	0.004 （0.05）	0.003 （0.65）	0.005 （1.13）
$Price_i$	$-4.401E-6^{***}$ （-7.35）	$-9.024E-6^{***}$ （-14.50）	$-8.180E-6^{***}$ （-10.81）	-1.431E-6 （-1.34）	-3.932E-7 （-0.36）	1.305E-6 （1.21）	$-1.248E-6^{**}$ （-3.70）	$-1.493E-5^{***}$ （-5.56）	$-2.561E-6$ （-0.70）
$Year_i$	0.001 （1.50）	0.001^{*} （1.70）	0.001^{**} （2.46）	0.001 （1.02）	0.001^{*} （1.86）	0.000 （0.62）	0.005^{***} （3.48）	0.005^{***} （3.97）	0.005^{***} （3.67）
Adjusted R^2	0.168	0.191	0.212	0.101	0.081	0.141	0.330	0.266	0.349
Number of observations	2329	2329	2329	1547	1548	1548	461	483	458

Table 3. Validation of Different Models in Hangzhou, Wuhan and Nanning: Based on the Jobs-housing Matching Rate of Workers. Notes: ***,**, and * denote significance levels of 0.001, 0.05, and 0.1, respectively; t-statistics are in parentheses.

Spatial econometric tests of OLS regression analysis

Spatial autocorrelation in both the independent and dependent variables necessitate the use of spatial econometric models, contingent upon an appropriate spatial weight matrix. Based on Queen's first-order contiguity, the global Moran's I value for OLS residuals of worker and resident jobs–housing matching rates were (0.266, 0.232) for Hangzhou, (0.444, 0.214) for Wuhan, and (0.253, 0.122) for Nanning, respectively (all $p < 0.05$). These results confirm significant positive spatial autocorrelation, justifying the application of spatial model correction.

Lagrange Multiplier (LM) tests were then conducted to determine the specific model form (Table 4). In Hangzhou, both SLM and SEM statistics were significant for the worker and resident models, with SEM

showing stronger significance; results were consistent across alternative weight matrices (second-order contiguity, distance, inverse distance). Wuhan exhibited a similar pattern, with SEM preferred. In Nanning, both models were significant, but SLM was slightly stronger, consistent across weight matrices, therefore SLM was adopted. These model choices align with city characteristics: Hangzhou and Wuhan, as larger provincial capitals, exhibit spatial effects primarily through unobserved factor transmission (spatial error), while Nanning, as a smaller capital, shows direct neighborhood spillover effects (spatial lag).

Variable	Hangzhou		Wuhan		Nanning	
	Worker JHM	Resident JHM	Worker JHM	Resident JHM	Worker JHM	Resident JHM
Constant	0.859*** (27.48)	0.851*** (21.85)	0.662*** (22.85)	0.809*** (29.38)	0.522*** (6.25)	0.359*** (3.89)
$D_{main,i}$	8.912E-7 (1.06)	9.659E-7 (0.92)			1.841E-5*** (4.12)	1.029E-5** (2.09)
$D_{sub,i}$	-1.414E-6* (-1.84)	7.657E-7 (0.80)	2.491E-6** (2.00)	2.295E-6* (1.94)		
$D_{fensan,i}$	-6.798E-8 (-0.04)	1.113E-6 (0.51)	6.054E-6** (2.74)	2.978E-6 (1.42)	-1.609E-5* (-1.75)	-1.517E-5 (-1.49)
$D_{cen,i}$	−0.007*** (-3.16)	−0.012*** (-4.30)	−0.015*** (-5.17)	−0.018*** (-6.58)	-8.531E-6 (-0.83)	1.743E-5 (1.53)
$HHI_{cen,i}$	−0.082*** (-4.52)	0.078*** (3.47)	0.037* (1.95)	-0.020 (-1.10)	-0.012 (0.52)	0.055** (2.11)
$Spe_{cen,i}$	0.000 (-0.05)	−0.007*** (-2.72)	8.989E-5 (0.21)	-0.0003 (-0.63)	−0.007** (-2.69)	0.0005 (0.20)
$Emp_{cen,i}$	3.441E-7* (1.66)	3.029E-7 (1.18)	-1.488E-6** (-2.39)	1.638E-6** (2.78)	-3.120E-6 (-1.38)	6.434E-6** (2.57)
$Den_{cen,i}$	-1.236E-5*** (3.12)	4.420E-6 (0.90)	1.008E-6 (0.31)	4.221E-6 (1.37)	1.053E-5* (1.72)	-1.137E-5* (-1.68)
$E_{Rsqu_{cen,i}}$	−0.203** (-2.26)	0.004 (0.97)	−0.656*** (-4.48)	0.385** (2.77)	-0.060 (-0.62)	-0.137 (-1.29)
$H_{Rsqu_{cen,i}}$	0.006 (0.12)	−0.153** (-2.47)	0.149 (1.44)	−0.339*** (-3.43)	0.015 (0.49)	-0.003 (-0.08)
$ENT_{i,j}$	0.161*** (4.88)	0.034 (0.83)	-0.050 (-1.41)	0.040（1.19）	-0.098 (-1.26)	0.234** (2.72)
BAL_i	0.051*** (3.40)	−0.076*** (-4.08)	0.089*** (4.80)	−0.071*** (-4.07)	0.055 (1.00)	0.137** (2.28)
Pub_i			0.001 (1.01)	-0.000 (-0.16)		
Edu_i	-0.001 (-1.29)	0.001** (2.67)	0.001*（1.15）	0.001 (1.64)	0.014*** (4.46)	0.007** (2.01)
$D_{metro,i}$	-1.422E-5 (-0.03)	−0.003** (-4.28)	-0.002 (-1.39)	−0.006*** (-4.61)	-0.005* (-1.80)	-0.007** (-2.13)
$D_{huanxian,i}$	−0.003*** (-2.69)	0.002 (1.58)	0.002* (1.71)	0.005** (4.35)	−0.007* (-1.84)	0.015*** (3.37)
$D_{infra,\ i}$	−0.005*** (-4.55)	0.001 (0.87)	-0.001 (-0.55)	-0.002 (-1.01)	0.008* (1.95)	-0.001 (-0.25)
$D_{hinder,i}$	0.006*** (7.39)	0.003** (2.53)	0.006** (3.06)	0.005** (3.10)	0.006 (1.31)	0.015*** (3.31)
$Price_i$	-8.180E-6*** (-10.81)	-1.958E-6** (-2.08)	1.823E-7 (0.17)	-8.210E-7 (-0.80)	-2.916E-6 (-0.79)	-2.195E-6 (-0.54)
$Year_i$	0.001** (2.46)	-0.004*** (-5.13)	0.0002 (0.30)	-0.003*** (-4.14)	0.004** (3.27)	-0.005*** (-3.58)
Adjusted R^2	0.212	0.102	0.121	0.131	0.345	0.103
Number of observations	2329	2329	1548	1548	458	458
F-statistic Prob(F-statistic)	34.06 (0.0000)	14.89 (1.40e-045)	12.18 (1.63e-035)	13.23 (5.24e-039)	14.35 (4.56e-034)	3.93 (1.60e-007)
Akaike info criterion	-3215.56	-2191.59	-1454.06	-1615	-653.962	-563.423
Koenker (BP) statistic	115.31 (0.0000)	57.84 (0.0001)	116.11 (0.0000)	62.02 (0.0000)	51.50 (0.0004)	34.61 (0.01)
Jarque-Bera statistic	1108.71 (0.0000)	5567.53 (0.0000)	355.23 (0.0000)	3300.95 (0.0000)	92.34 (0.0000)	1962.80 (0.000)

Moran's I(residuals)	0.266 (0.0000)	0.232 (0.0000)	0.444 (0.0000)	0.214 (0.0000)	0.253 (0.0000)	0.122 (0.0000)
Log LIK	1627.78	1115.79	747.03	827.50	345.98	300.71
LM-Lag	266.06	254.40	271.52	81.14	71.36	18.20
Robust LM-Lag	8.20	7.63	7.61	8.82	14.81	4.40
LM-Error	400.68	312.34	530.28	124.12	70.38	16.78
Robust LM-Error	142.82	74.57	266.37	51.80	13.83	2.98

Table 4. Spatial Econometric Tests of OLS Regression for Factors Influencing Jobs–Housing Matching. JHM is the abbreviation for Jobs–Housing Matching. Note: ***,**, and * denote significance levels of 0.001, 0.05, and 0.1, respectively; t-statistics are in parentheses.

Spatial Regression (SEM/SLM) Model Analysis

Comparison between OLS and spatial regression (SEM/SLM) models

Spatial regression models outperformed OLS across all three cities, as evidenced by higher Log-likelihood values, lower Akaike Information Criterion (AIC) values, and residual Moran's I values approaching zero after correction (Table 5). The R^2 improved substantially, with the largest increase observed for Wuhan's worker model, followed by Hangzhou (increases of 0.1271 and 0.1249 for workers and residents, respectively), and the smallest increase for Nanning's resident model (+0.0713). Worker models consistently yielded higher R^2 than resident models, suggesting greater unexplained variance in resident matching due to unobserved factors such as preferences and household characteristics. Nanning's highest worker model fit may reflect its smaller urban size and simpler spatial variation.

Spatial coefficients were highly significant across all models. For Hangzhou, the SEM lambda values were 0.443 (workers, z=17.39) and 0.421 (residents, z=16.17), indicating that approximately 42–44% of the spatial error was transmitted across neighboring regions. Wuhan exhibited higher lambda values of 0.566 (workers, z=24.67) and 0.315 (residents, z=10.28), reflecting stronger spatial dependence, especially for workers. Nanning, the SLM spatial lag coefficients were 0.339 (workers, z=7.98) and 0.185 (residents, z=3.99), indicating that a one-unit increase in neighboring matching rates raises local rates by 0.339 and 0.185 units, respectively, with stronger spillover effects observed for workers.

Core variable signs remained consistent between the OLS and spatial models, confirming the robustness of the findings and the stability of the underlying economic-geographic logic.

Variable	Hangzhou（SEM）		Wuhan（SEM）		Nanning（SLM）	
	Worker JHM	Resident JHM	Worker JHM	Resident JHM	Worker JHM	Resident JHM
Constant	0.879***	0.844***	0.688***	0.818***	0.253***	0.227**
	（19.48）	（15.26）	（17.75）	（24.46）	（3.08）	（2.43）
$D_{main,i}$	-1.250E-7	1.350E-6			1.390E-5**	9.865E-6**
	（-0.10）	（0.86）			（3.08）	（2.09）
$D_{sub,i}$	-1.498E-6	1.021E-6	2.898E-8	1.831E-6		
	（-1.24）	（0.70）	（0.02）	（1.21）		
$D_{fensan,i}$	-1.292E-6	1.930E-7	4.616E-6	2.891E-6	-6.661E-6	-1.040E-5
	(-0.49)	(0.06)	(1.36)	(1.07)	(-0.79)	(-1.06)
$D_{cen,i}$	−0.006*	−0.013***	−0.011**	−0.017***	-5.480E-6	1.486E-5（1.36）
	（-1.88）	（-3.29）	（-2.58）	（-5.15）	（-0.59）	
$HHI_{cen,i}$	−0.053**	0.054**	0.031	-0.012	-0.011	0.047*
	（-2.58）	（2.11）	（1.60）	（-0.64）	（-0.51）	（1.87）
$Spe_{cen,i}$	-4.431E-5	−0.006**	-0.0001	-0.0002	-0.005**	0.0004
	（0.99）	（-1.98）	（-0.24）	（-0.43）	（-2.21）	（0.15）
$Emp_{cen,i}$	1.714E-7	8.615E-8	-8.226E-7	1.150E-6*	-2.543E-6	5.419E-6**
	（0.79）	（0.32）	（-1.30）	（1.80）	（-1.24）	（2.25）
$Den_{cen,i}$	7.319E-6	1.821E-6	-8.702E-7	2.594E-6	7.523E-6	-9.663E-6
	（1.73）	（0.34）	（-0.25）	（0.77）	（1.36）	（-1.49）
$E_{Rsqu_{cen,i}}$	−0.116	-0.039	−0.243*	0.241	-0.030	-0.115
	（-1.17）	（-0.31）	（-1.67）	（1.63）	（-0.35）	（-1.13）
$H_{Rsqu_{cen,i}}$	0.011	-0.052	0.036	−0.217**	0.013	-0.0015
	（0.19）	（-0.74）	（0.34）	（-2.08）	（0.46）	（-0.05）
$ENT_{i,j}$	0.169***	0.041	-0.090*	0.011	-0.087	0.187**
	（3.72）	（0.73）	（-1.93）	（0.26）	（-1.23）	（2.25）
BAL_i	0.040*	−0.072**	0.045*	−0.065**	0.028	0.125**
	（1.87）	（-2.73）	（1.72）	（-2.95）	（0.57）	（2.17）
Pub_i			0.0003	-0.001		
			（0.21）	（-0.13）		

Edu_i	-0.0002 （-0.25）	0.001 （1.54）	0.001 （1.17）	0.001 （0.96）	0.008** （2.73）	0.006* （1.64）
$D_{metro,i}$	0.0001 （0.13）	−0.003** （-3.15）	−0.004** （-1.99）	−0.006*** （-3.46）	-0.002 （0.64）	-0.005 （-1.47）
$D_{huanxian,i}$	-0.002 （-1.37）	0.002 （1.12）	0.005** （2.19）	0.005*** （3.17）	-0.003 （-0.92）	0.014** （3.28）
$D_{infra,\ i}$	−0.004** （-2.21）	-0.0004 （-0.18）	0.002 （0.61）	-0.002 （-0.87）	0.005 （1.43）	-0.001 （-0.31）
$D_{hinder,i}$	0.006*** （4.84）	0.003* （1.89）	0.007** （2.48）	0.006** （2.63）	0.007* （1.94）	0.015*** （3.42）
$Price_i$	-8.348E-6*** （-7.67）	-2.174E-6 （-1.63）	1.561E-6 （1.37）	-2.477E-7 （-0.22）	1.394E-6 （0.41）	3.377E-6 （0.87）
$Year_i$	0.0006 （0.95）	-0.002** （-3.16）	-0.0003 （0.35）	-0.003*** （-3.53）	0.005*** （4.38）	−0.004** （-2.52）
W_w_bal					0.339*** （7.98）	
W_h_bal						0.185*** （3.99）
λ Lambda	0.443*** （17.39）	0.421*** （16.17）	0.566*** （24.67）	0.315*** （10.28）		
R^2	0.3391	0.2269	0.4040	0.2170	0.4621	0.1743
Number of observations	2329	2329	1548	1548	458	458
Log LIK	1773.23	1237.04	965.14	879.09	376.74	308.88
Akaike info criterion	-3506.45	-2434.08	-1890.27	-1718.18	-713.479	-577.769
Moran's I(residuals)	-0.040	-0.025	-0.060	-0.029	0.030	0.018

Table 5. Spatial Regression (SEM/SLM) Analysis of Factors Influencing Jobs–Housing Matching. Notes: ***,**, and * denote significance levels of 0.001, 0.05, and 0.1, respectively; t-statistics are in parentheses. W_w_bal and W_h_bal represent the spatial lag variables of the dependent variables—worker and resident jobs–housing matching rate, respectively.

Urban Spatial Structure Factors

Among the polycentric variables, distance to the nearest employment center ($D_{cen,i}$) emerges as the most consistently significant determinant. In Hangzhou and Wuhan, $D_{cen,i}$ exerts a significant negative effect on both worker and resident jobs–housing matching rates (Hangzhou: p<0.1 and p< 0.001; Wuhan: p<0.05 and p<0.001), confirming the classic distance decay effect in urban economics. Notably, the absolute coefficient values for residents are consistently larger than those for workers (Hangzhou: -0.013*** vs. -0.006*; Wuhan: -0.017*** vs. -0.011**), indicating that residents are more sensitive to employment center proximity. This asymmetry echoes earlier supportive studies that found employment polycentricity shortened commute times for workers in Los Angeles [2] and improved job accessibility in the Dutch Randstad [4]. However, the insignificant effects of $D_{main,i}$ *and* $D_{sub,i}$ in Hangzhou and Wuhan suggest that, after controlling for distance to the nearest center, the independent effects of primary and subcenters are absorbed, reflecting mature but uneven polycentric structures where subcenters remain underdeveloped-a pattern contrasting with findings in Seoul [13] and Beijing [12]. In Nanning, $D_{main,i}$ is significantly positive for both workers and residents (p<0.0.05), implying that primary center has reached saturation and peripheral new districts achieve better jobs–housing balance, a pattern consistent with the limited applicability of the co-location hypothesis observed in French cities[11]. The non-significance of $D_{fensan,i}$ across all three cities suggests that dispersed employment points have negligible impacts, corroborating evidence that suburban employment centers often exhibit low self-containment and high vehicle miles traveled[21–23]. Together, these results underscore that the effect of polycentric structure on jobs–housing matching is highly context-dependent, varying by city size, center maturity, and spatial configuration.

Regarding the impact of employment center industrial structure, the Hirschman–Herfindahl Index ($HHI_{cen,i}$) shows a significant negative effect on worker matching (p<0.0.05) and a significant positive effect on resident matching (p<0.05) in Hangzhou. This suggests that industrial diversification benefits residents by providing diverse employment opportunities that facilitate nearby job access, yet for workers, diversification may imply insufficient employment agglomeration, thereby reducing matching efficiency. In Wuhan and Nanning, the effects of industrial diversification are insignificant, except for a marginally significant positive effect in Nanning resident model (0.047*). The relative specialization index ($Spe_{cen,i}$) exhibits significant negative effects in Hangzhou's resident model (-0.006**) and Nanning's worker model (-0.005*), indicating that excessive industrial specialization reduces jobs–housing matching levels, as overly specialized employment centers offer a narrow range of job types that meet diverse employment needs.

Regarding the impact of polycentric development level, employment center size ($Emp_{cen,i}$) and density ($Den_{cen,i}$) are not significant in most models, except for the resident models in Wuhan (p<0.0.1) and Nanning (p<0.0.05), where both show significantly positive effects, indicating that larger employment centers favor residents' jobs–housing matching. However, this contrasts with evidence from Atlanta, where larger employment centers were associated with longer commute times [24] , suggesting that the effect of center size is

contingent on urban context and population group. The functional polycentricity index ($E_{Rsqu_{cen,i}}$) is significantly negative in the Wuhan's worker model (-0.243*), suggesting that a higher degree of functional polycentricity- implying greater spatial separation of employment and residential functions, thereby increasing commuting distances and mismatch risk. Similarly, the morphological polycentricity index ($H_{Rsqu_{cen,i}}$) is significantly negative in the Wuhan's resident model (-0.217**), indicating adverse effects on residents' matching. This aligns with seven Spanish urban systems, polycentricity reduced commuting distance, especially with equitable employment distribution[16]. Overall, functional and morphological polycentricity exert differential effects across population groups. Our results across Hangzhou, Wuhan, and Nanning confirm that the impact of polycentric development is highly context-dependent, shaped by configuration, maturity, and population segment.

Neighborhood Built Environment Factors

The impact of land use mix ($ENT_{i,j}$) exhibits pronounced city-specific differences. In Hangzhou's worker model, it is significantly positive at the 0.001 level (0.169***), indicating that diversified land use provides more nearby employment opportunities and improves work matching- consistent with prior evidence that higher land use mix reduces commuting distance [38]. In Wuhan's worker model, however, it is significantly negative at the 0.1 level (-0.090*), possibly because mixed land use in Wuhan's spatial configuration leads to declining employment density and jobs per resident [14]. In Nanning's resident model, it is significantly positive at the 0.05 level (0.187**), suggesting that land use mix benefits residents in smaller cities.

The balance between residential and non-residential land use (BAL_i) is one of the most consistently significant neighborhood variables, appearing in five of the six models with a clear role-based divergence. It is consistently positive for worker jobs–housing matching (Hangzhou 0.040*, Wuhan 0.045*, Nanning 0.028 n.s.), indicating that balanced land use facilitates workers' nearby employment. Conversely, it is significantly negative for resident jobs–housing matching in Hangzhou (-0.072**) and Wuhan (-0.065**),suggesting that balancing land use toward non-residential uses may crowd out residential space and reduce resident matching levels. Nanning's positive resident effect (0.125**) deviates from this pattern, likely due to its smaller city size and different land use dynamics.

Educational facility density (Edu_i) is significant only in Nanning, with positive effects on both workers (0.008**) and residents (0.006*), suggesting that adequate educational services improve jobs–housing matching by influencing residential location choices. In Hangzhou and Wuhan, the effect is not significant, possibly because educational resource allocation is relatively balanced, yielding insufficient variation. Public facility density (Pub_i) appears only in the Wuhan's model and is not significant, indicating limited direct impact–consistent with evidence that public service provision moderates commuting patterns primarily through indirect channels such as housing type diversity and transit capacity [16] .

Transportation Accessibility Factors

Metro accessibility ($D_{metro,i}$) is significantly negative in Hangzhou's resident model (-0.003**, p<0.0.05) and Wuhan's resident (-0.006***, p<0.0.001) and worker models (--0.004**, p<0.0.05), indicating that proximity to metro stations improves jobs–housing matching, with stronger effects for residents than workers. In Nanning, metro accessibility is not significant, possibly due to limited metro coverage (one line vs. three in Hangzhou and five in Wuhan).

Distance to the outer ring road ($D_{huanxian,i}$) is significantly positive in Wuhan's resident (0.005***, p<0.0.001) and worker models (0.005**, p<0.0.05), and in Nanning's resident model (0.014**, p<0.0.05), suggesting that locations closer to the ring road experience worse matching. This reflects the ring road's role as a spatial barrier in concentrically expanding cities (Wuhan, Nanning). Hangzhou, with more balanced intra-urban development, shows no significant effect. This difference reflects each city's spatial structure: Nanning's construction is concentrated inside the ring road, making it a strong barrier; Wuhan's concentric expansion gives the ring road a prominent separating effect; Hangzhou's balanced development inside and outside its single ring road renders the barrier effect insignificant.

Distance to major infrastructure ($D_{infra,\ i}$) is significantly negative only in Hangzhou's worker model (-0.004**, p<0.0.05), indicating that greater distance from major infrastructure reduces the worker jobs–housing matching rate. This likely reflects the concentration of employment opportunities near such facilities, which facilitates nearby employment for workers. In other cities and models, this variable is not significant, suggesting that the impact of major infrastructure on jobs–housing matching varies by city.

Natural barrier distance ($D_{hinder,i}$) is the most consistently significant control variable, positive across all six models (coefficients: 0.003* to 0.015***), confirming that areas farther from natural barrier achieve higher matching rates. The effect is strongest in Nanning (mountainous terrain), moderate in Wuhan (dense river network), and weakest in Hangzhou (relatively flat topography), supporting the classic view that natural barriers disrupt intra-urban functional connectivity.

Urban Economic Development Factors

Housing price ($Price_i$) is significantly negative only in the Hangzhou's worker model (-8.348E-6***, z = -7.67, p<0.0.001), representing one of the strongest effects in the full model. This indicates that the higher housing prices reduce worker jobs–housing matching rates, reflecting the crowding-out effect of high housing costs that drives jobs–housing separation. This finding aligns with Levine's study in Minneapolis [29]. At the time of the study, Hangzhou had the highest mean housing price (23,703 yuan/m²) and largest standard deviation (8,585), whereas Wuhan and Nanning had lower means and smaller dispersions (14,542 yuan/m², 4,718; 10,266 yuan/m², 2,596, respectively), explaining the non-significant results in those cities. This suggests that in cities with greater spatial housing price variation, high-price areas are more prone to jobs–housing imbalance, reinforcing the policy implication that housing affordability interventions should prioritize workers in high-cost urban contexts [35].

Building age ($Year_i$) exhibits clear city-specific differences. In Hangzhou's and Wuhan's resident models of, it is significantly negative (-0.002** and -0.003***, respectively), indicating that longer building ages are associated with lower resident matching rates. This likely reflects the gradual decoupling of long-term residents' residences from their evolving workplaces, suggesting that urban renewal should introduce employment functions into aging areas. In Nanning, however, building age is significantly positive for workers (0.005***) and significantly negative for residents (-0.004**). This bidirectional effect reflects Nanning's developmental stage: its urban economy and public services remain concentrated in older districts, which continue to attract workers and facilitate on-site matching, while new districts remain underdeveloped. This finding resonates with et al.'s study of Guangzhou and Shenzhen [35].

Discussion

Re-examination of Core Influencing Factors: Robustness and Pseudo-Regression under Spatial Effects

First, spatial dependence fundamentally shapes jobs–housing matching. OLS models that neglect spatial autocorrelation suffer from serious specification bias: Moran's I value of OLS residuals are highly significant across all three cities, and introducing spatial effects substantially improves model fit—most notably in Wuhan's worker model (R^2 increase of 233.9%). The three cities exhibit fundamentally different spatial dependence mechanisms, reflecting their distinct developmental stages. Hangzhou and Wuhan adopted the spatial error model (SEM), indicating that spatial autocorrelation is primarily transmitted through unobserved spatial factors in neighboring areas (e.g., regional policies, environmental quality, historical planning inertia). Wuhan's lambda reaches 0.566 (z = 24.67), the highest among all models, reflecting its unique three-town structure divided by the Yangtze and Han Rivers, where limited cross-river corridors amplify spatial dependence. Hangzhou's lambda (0.42–0.44) indicates moderate spatial dependence, consistent with its mature polycentric structure where employment centers maintain both independence and connectivity. In contrast, Nanning adopts the spatial lag model (SLM), with significant spatial spillover effects (W_w_bal = 0.339, W_h_bal = 0.185), reflecting its smaller size and more compact spatial interactions where neighboring matching rates directly influence local outcomes.

Second, comparing OLS and spatial econometric models reveals a "pseudo-regression" phenomenon. Several variables significant in OLS models lost significance or reversed direction after controlling for spatial effects. For instance, $D_{sub,i}$ in Wuhan is significant in OLS but becomes insignificant in SEM, indicating that subcenter effects are masked by the primary center's spatial radiation. $Den_{cen,i}$ in Hangzhou switched from negative (OLS) to positive (SEM), cautioning that without controlling for spatial errors, one might erroneously conclude that density inhibits matching when the opposite may be true. These findings underscore that spatial econometric models are not merely a technical correction but a necessary lens for uncovering the true mechanisms underlying jobs–housing relationships across cities at different developmental stages.

Comparative Heterogeneity of Urban Development Models: Polycentric Polarization, Uneven Polycentricity, and Polycentric Cluster Balance

Hangzhou's polycentric polarization model is characterized by a core contradiction of center dependence superimposed with functional singularity. The significant negative effect of $D_{cen,i}$ confirms that the polycentric structure improves matching by dispersing employment opportunities. However, industrial diversification (($HHI_{cen,i}$) exerts opposing effects on workers (negative) and residents (positive), reflecting the complexity of a highly developed industrial spatial structure. The negative worker effect points to functional specialization in core areas (e.g., purely commercial CBDs) exacerbating jobs–housing separation. Extremely high housing prices ($Price_i$) produce a significant crowding-out effect, entrenching this separation and reflecting the systemic dilemma of high-cost cities. The selection of SEM (lambda ≈ 0.42–0.44) suggests that this issue is profoundly influenced by macro-structural factors such as historical planning and regional policies.

Wuhan's uneven polycentric model is characterized by extremely spatial dependence and a complex ring-road/corridor barrier effect. The negative effect of $D_{cen,i}$ conforms to distance decay, confirming that polycentricity improves matching. Wuhan's exceptionally high spatial error coefficient (lambda=0.566, the highest among all models) indicates that cross-regional spillovers are key to understanding its jobs-housing dynamics; policies targeting a single area may generate significant spatial externalities. The positive significance of $D_{huanxian,i}$ confirms the ring road as a substantive spatial barrier, while the significance role of highlights rail transit's critical function in connecting river- and ring road-segmented urban clusters. However, Wuhan's polycentric development exhibits notable unevenness: subcenter variables are weak, and functional polycentricity (($E_{Rsqu_{cen,i}}$) negatively affects worker matching, suggesting that its polycentric structure remains in an early stage of "morphological dispersion" rather than "functional complementarity." As a traditional industrial city with clear functional zoning (e.g., Optics Valley, Auto Valley), increased functional polycentricity may further concentrate employment in specific areas, exacerbating spatial imbalance.

Nanning's polycentric cluster balance embodies e$D_{metro,i}$ndogenous balance in peripheral clusters and functional saturation of the main center. Contrary to Hangzhou and Wuhan, the positive effect of $D_{main,i}$ indicates that in small cities, the main-center's matching capacity has reached saturation, while peripheral new development zones achieve better balance—a pattern resembling "reverse commuting" in Western studies. The selection of SLM and significant spatial lags coefficients indicate that Nanning's success relies on positive spillovers between built-up areas, with employment opportunity diffusion being a key driver. The significant positive effect of Edu_i further demonstrates that high-quality public services are an important factor in attracting population settlement and promoting local employment during rapid development.

Policy Implications: From Spatial Intervention to City-Specific Governance

First, governance should shift from plot-level management to spatial collaborative governance. Due to strong spatial dependence and spillover effects identified in this study, any jobs–housing balance policy targeting a single area may generate spatial externalities affecting surrounding areas. For cities with spatial error transmission patterns like Hangzhou and Wuhan (Lambda=0.42-0.44 and 0.566, respectively), priority should be given to coordinating unobserved spatial factors—regional industrial policies, environmental quality, and social service allocation. For cities dominated by spatial spillover effects like Nanning (W_w_bal=0.339, W_h_bal=0.185), a "point-driven-area" strategy should be adopted, leveraging positive spillovers from successful clusters to drive improvements in surrounding areas.

Second, planning strategies must be city-specific. For Hangzhou, the policy focus should be on promoting industrial diversification to counteract the functional singularity of core areas, while expanding affordable and rental housing supply to mitigate the crowding-out effect of high housing prices. For Wuhan, efforts should prioritize densifying rail transit networks across rivers and ring roads (given the significant barrier effect of $D_{huanxian,i}$ and the mitigating role of $D_{metro,i}$), and cultivating subcenters like Optics Valley and Auto Valley into functionally complete urban nodes rather than mere industrial bases, addressing the negative effect of functional polycentricity. For Nanning, the key lies in consolidating successful peripheral clusters (benefiting from positive $D_{main,i}$) while revitalizing the saturated primary center through micro-renewal, and leveraging the significant positive effect of educational facilities to attract population and promote local employment.

Third, systemic coordination of polycentric policies with land, transportation, and housing policies is essential. Employment center siting should be optimized based on residential distribution and commuting costs to shorten commutes at the source. Refined land use regulation should ensure a guaranteed proportion of affordable residential space near employment centers, balancing the dual effects of land use mix (positive for workers, negative for residents in larger cities). Finally, transit-oriented housing policies—such as mandating affordable housing supply near metro stations—should be implemented to ensure that the positive externalities of public transportation benefit all income groups, addressing the limited effect of metro accessibility on resident matching ($D_{metro,i}$ significant but with small coefficients).

Conclusion

This study compares OLS and spatial econometric models across three Chinese cities (Hangzhou, Wuhan, Nanning) to examine jobs–housing matching determinants. Four key findings emerge. First, spatial dependence is indispensable for analyzing jobs–housing matching. OLS residuals exhibit significant spatial autocorrelation across all three cities; spatial econometric models substantially improve model fit (R^2) and reduce AIC values. The three cities exhibit fundamentally different spatial dependence mechanisms: Hangzhou and Wuhan follow the spatial error model (SEM), indicating that unobserved factors (e.g., regional policies, environmental quality) drive spatial effects, while Nanning follows the spatial lag model (SLM), where neighboring matching rates directly diffuse to local areas. Second, after controlling for spatial effects, influencing factors varied significantly by city size and developmental stage. Distance to the nearest employment center showed consistent negative effects in Hangzhou and Wuhan, validating the classic distance decay law. However, distance to the primary center in Nanning is significantly positive, reflecting main center saturation and better balance in peripheral new districts—confirming that polycentric effects are moderated by urban developmental stage. Third, land use and transportation effects exhibit complexity. Land use mix generally benefits worker matching but negatively affects resident matching in larger cities, warning that excessive functional mixing may crowd out residential space. Metro accessibility shows limited or negative effects on resident matching across all cities, suggesting transit capitalization effects may raise housing costs and displace residents—underscoring the need to distinguish worker and resident perspectives. Fourth, cities at different stages exhibit distinct dominant logics. Hangzhou shows polycentric polarization with high-price crowding-out; Wuhan displays strong spatial dependence with ring-road/river barriers and immature subcenters; Nanning demonstrates peripheral cluster balance with positive spillovers and main-center saturation.

The theoretical contributions are twofold: (1) empirically demonstrating the severity of neglecting spatial effects in urban spatial analysis through rigorous model comparison; (2) constructing and validating a comprehensive analytical framework integrating polycentric structure, built environment, transportation accessibility, and control variables. Limitations include the use of cross-sectional data (future studies should adopt panel data and spatiotemporal models), the inability to capture micro-level heterogeneity (e.g., income, family structure), and the limited city sample (future research should expand to more city types to test generalizability). Achieving high-level jobs–housing matching requires respecting urban developmental stages, leveraging spatial laws, and coordinating multi-dimensional policies.

Data Availability

Mobile signaling data used for identifying commuting populations were obtained from China Mobile and are subject to confidentiality agreements; therefore, these data are not publicly available. OpenStreetMap (OSM) data are freely available at https://www.openstreetmap.org. Points of Interest (POI) data were sourced from the Amap (Gaode) API and are accessible via https://lbs.amap.com. The national land use classification dataset developed by Gong et al. [48] and the urban construction land temporal sequence dataset by Gong et al. [49] are available from the corresponding author upon reasonable request. Residential community housing price data were collected from SouFun.com (https://www.fang.com) via web scraping; these data are publicly accessible but subject to the website's terms of use. Economic census data for Hangzhou were obtained from the local statistical bureau and are not publicly available due to confidentiality restrictions.

References

1.Gordon, P., Richardson, H. W. & Jun, M.-J. The Commuting Paradox Evidence from the Top Twenty. Journal of the American Planning Association 57, 416–420 (1991).
2.Giuliano, G. & Small, K. A. Is the Journey to Work Explained by Urban Structure? Urban Studies 30, 1485–1500 (1993).

3.Modarres, A. Polycentricity, Commuting Pattern, Urban Form: The Case of Southern California: Polycentric commuting patterns in southern California. International Journal of Urban and Regional Research 35, 1193–1211 (2011).
4.Van Ham, M., Hooimeijer, P. & Mulder, C. H. Urban Form and Job Access: Disparate Realities in the Randstad. Tijd voor Econ & Soc Geog 92, 231–246 (2001).
5.Alpkokin, P., Cheung, C., Black, J. & Hayashi, Y. Dynamics of clustered employment growth and its impacts on commuting patterns in rapidly developing cities. Transportation Research Part A: Policy and Practice 42, 427–444 (2008).
6.Loo, B. P. Y. & Chow, A. S. Y. Jobs-housing balance in an era of population decentralization: An analytical framework and a case study. Journal of Transport Geography 19, 552–562 (2011).
7.Zhao, P., Lü, B. & Roo, G. D. Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. Journal of Transport Geography 19, 59–69 (2011).
8.Sun, B., He, Z., Zhang, T. & Wang, R. Urban spatial structure and commute duration: An empirical study of China. International Journal of Sustainable Transportation 10, 638–644 (2016).
9.Zhu, J., Niu, X. & Shi, C. The Influencing Factors of a Polycentric Employment System on Jobs-Housing Matching—A Case Study of Hangzhou, China. Sustainability 11, 5752 (2019).
10.Veneri, P. Urban Polycentricity and the Costs of Commuting: Evidence from Italian Metropolitan Areas. Growth and Change 41, 403–429 (2010).
11.Aguilera, A. Growth in Commuting Distances in French Polycentric Metropolitan Areas: Paris, Lyon and Marseille. Urban Studies 42, 1537–1547 (2005).
12.Hu, L., Sun, T. & Wang, L. Evolving urban spatial structure and commuting patterns: A case study of Beijing, China. Transportation Research Part D: Transport and Environment 59, 11–22 (2018).
13.Jun, M.-J. The effects of polycentric evolution on commute times in a polycentric compact city: A case of the Seoul Metropolitan Area. Cities 98, 102587 (2020).
14.Schwanen, T., Dieleman, F. M. & Dijst, M. The Impact of Metropolitan Structure on Commute Behavior in the Netherlands: A Multilevel Approach. Growth and Change 35, 304–333 (2004).
15.Sultana, S. & Weber, J. Journey-to-Work Patterns in the Age of Sprawl: Evidence from Two Midsize Southern Metropolitan Areas*. The Professional Geographer 59, 193–208 (2007).
16.Duarte, C. M. & Fernández, M. T. The Influence of Urban Structure on Commuting: An Analysis for the Main Metropolitan Systems in Spain. Procedia Engineering 198, 52–68 (2017).
17.Schwanen, T., Dieleman, F. M. & Dijst, M. Car Use in Netherlands Daily Urban Systems: Does Polycentrism Result in Lower Commute Times?1. Urban Geography 24, 410–430 (2003).
18.Angel, S. & Blei, A. M. The spatial structure of American cities: The great majority of workplaces are no longer in CBDs, employment sub-centers, or live-work communities. Cities 51, 21–35 (2016).
19.Zhao, P., Lu, B. & De Roo, G. The impact of urban growth on commuting patterns in a restructuring city: Evidence from Beijing. Papers in Regional Science 90, 735–755 (2011).
20.Asikhia, M. O. & Nkeki, N. F. Polycentric Employment Growth and the Commuting Behaviour in Benin Metropolitan Region, Nigeria. JGG 5, p1 (2013).
21.Cervero, R. & Wu, K.-L. Polycentrism, Commuting, and Residential Location in the San Francisco Bay Area. Environ Plan A 29, 865–886 (1997).
22.Cervero, R. & Wu, K.-L. Sub-centring and Commuting: Evidence from the San Francisco Bay Area, 1980-90. Urban Studies 35, 1059–1076 (1998).
23.Sultana, S. Job/Housing Imbalance and Commuting Time in the Atlanta Metropolitan Area: Exploration of Causes of Longer Commuting Time. Urban Geography 23, 728–749 (2002).
24.Sultana, S. Some Effects of Employment Centers on Commuting Times in the Atlanta Metropolitan Area, 1990. sgo 40, 225–233 (2000).
25.Schwanen, T., Dieleman, F. M. & Dijst, M. Travel behaviour in Dutch monocentric and policentric urban systems. Journal of Transport Geography 9, 173–186 (2001).
26.Yang, J., French, S., Holt, J. & Zhang, X. Measuring the Structure of U.S. Metropolitan Areas, 1970–2000: Spatial Statistical Metrics and an Application to Commuting Behavior. Journal of the American Planning Association 78, 197–209 (2012).
27.Wachs, M., Taylor, B. D., Levine, N. & Ong, P. The Changing Commute: A Case-study of the Jobs-Housing Relationship over Time. Urban Studies 30, 1711–1729 (1993).
28.Peng, Z.-R. The Jobs-Housing Balance and Urban Commuting. Urban Studies 34, 1215–1235 (1997).
29.Levine, J. Rethinking Accessibility and Jobs-Housing Balance. Journal of the American Planning Association 64, 133–149 (1998).
30.Lee, S., Seo, J. G. & Webster, C. The Decentralising Metropolis: Economic Diversity and Commuting in the US Suburbs. Urban Studies 43, 2525–2549 (2006).
31.Lin, D., Allan, A. & Cui, J. Sub-centres, socio-economic characteristics and commuting: a case study and its implications. International Journal of Urban Sciences 21, 147–171 (2017).
32.Zhou, L. et al. Identify social and job disparities in the relationship between job-housing balance and urban commuting using Baidu trajectory big data. Transactions in GIS 28, 258–277 (2024).
33.Cervero, R. Jobs-Housing Balancing and Regional Mobility. Journal of the American Planning Association 55, 136–150 (1989).
34.Xiao, W., Wei, D. & Li, H. Understanding jobs-housing imbalance in urban china: a case study of Shanghai. J. Transp. Land Use 14, 1–27 (2021).
35.Zhang, Y. & Luo, F. Exploring the relationship between jobs and housing in China's megacities: the case of Guangzhou and shenzhen. J. Asian Archit. Build. Eng. 24, 3312–3328 (2025).
36.Zhou, J., Wang, Y., Cao, G. & Wang, S. Jobs-housing balance and development zones in China: a case study of Suzhou Industry Park. Urban Geography 38, 363–380 (2017).
37.Ewing, R. & Cervero, R. Travel and the Built Environment: A Meta-Analysis. Journal of the American Planning Association 76, 265–294 (2010).
38.Cervero, R. Jobs-Housing Balance Revisited: Trends and Impacts in the San Francisco Bay Area. Journal of the American Planning Association 62, 492–511 (1996).
39.Liu, F. et al. Construction of the node—place—jobs-housing model: analysis of employment-residential ratio in subway station areas of shenzhen, china's highest construction density zone. PLOS One 20, e0337576 (2025).

40.Huang, D., Liu, Z. & Zhao, X. Monocentric or Polycentric? The Urban Spatial Structure of Employment in Beijing. Sustainability 7, 11632–11656 (2015).
41.Lv, Y., Zheng, X., Zhou, L. & Zhang, L. Decentralization and Polycentricity: Spatial Changes of Employment in Beijing Metropolitan Area, China. Sustainability 9, 1880 (2017).
42.Zhang, T., Sun, B., Li, W., Dan, B. & Wang, C. Polycentricity or dispersal? The spatial transformation of metropolitan Shanghai. Cities 95, 102352 (2019).
43.Liu, X., Derudder, B. & Wang, M. Polycentric urban development in China: A multi-scale analysis. Environment and Planning B: Urban Analytics and City Science 45, 953–972 (2018).
44.Sun, T. A longitudinal study of changes in intra-metropolitan employment concentration in Beijing: Decentralisation, reconcentration and polycentrification. Urban Studies 57, 748–765 (2020).
45.Parr, J. B. Spatial Definitions of the City: Four Perspectives. Urban Studies 44, 381–392 (2007).
46.Zhu, J., Niu, X. & Wang, Y. Polycentric Urban Spatial Structure Identification Based on Morphological and Functional Dimensions: Evidence from Three Chinese Cities. Sustainability 16, 2584 (2024).
47.Zhu, J. & Wang, Y. A sensitivity study on the measurement of urban polycentricity in chinese cities: center definition, indicator selection, and their interaction effects. PLOS One 21, e0350134 (2026).
48.Gong, P. et al. Mapping essential urban land use categories in China (EULUC-China): preliminary results for 2018. Science Bulletin 65, 182–187 (2020).
49.Gong, P., Li, X. & Zhang, W. 40-Year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing. Science Bulletin 64, 756–763 (2019).
50.Giuliano, G. & Small, K. A. Subcenters in the Los Angeles region. Regional Science and Urban Economics 21, 163–182 (1991).
51.Muñiz, I., Garcia-López, M. À. & Galindo, A. The Effect of Employment Sub-centres on Population Density in Barcelona. Urban Studies 45, 627–649 (2008).
52.Shearmur, R. G. & Coffey, W. J. Urban Employment Subcenters and Sectoral Clustering in Montreal: Complementary Approaches to the Study of Urban Form. Urban Geography 23, 103–130 (2002).
53.Wan, D., Zhao, L., Chong, X. & Cui, Y. The mechanism by which jobs–housing mismatch affects urban land resource allocation efficiency under external shocks: an excess commuting perspective. Land 15, 166 (2026).
54.Zhou, X., Yeh, A. G. O. & Yue, Y. Spatial variation of self-containment and jobs-housing balance in Shenzhen using cellphone big data. Journal of Transport Geography 68, 102–108 (2018).
55.Shearmur, R. Travel From Home: An Economic Geography of Commuting Distances in Montreal. Urban Geography 27, 330–359 (2006).
56.Yan, L., Wang, D., Zhang, S. & Xie, D. Evaluating the multi-scale patterns of jobs-residence balance and commuting time–cost using cellular signaling data: a case study in Shanghai. Transportation 46, 777–792 (2019).
57.Vasanen, A. Functional Polycentricity: Examining Metropolitan Spatial Structure through the Connectivity of Urban Sub-centres. Urban Studies 49, 3627–3644 (2012).
58.Song, Y., Merlin, L. & Rodriguez, D. Comparing measures of urban land use mix. Computers, Environment and Urban Systems 42, 1–13 (2013).

Author Contributions

J. Z. Conceptualization, methodology, software, and wrote the article, J. Z. and S. C. formal analysis, resources, visualization; J. Z., S. C. and W. W. data curation; Y. W. supervised and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by "2025 Jiangsu Province Education Planning Project, grant number B/2025/02/135" and "Start-up Fund for New Ph.D. Researchers of Suzhou Chien-Shiung Institute of Technology（60000001）".

Declarations

Competing interests

The authors declare no competing interests.