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 Patterns and Determinants of Industrial Operation in Maharashtra: A District-Level Spatial Econometric Analysis

Sudeb Pal

sudebcu@gmail.com

University of Calcutta https://orcid.org/0000-0001-9165-3661

Sumana Bandyopadhyay

University of Calcutta

Research Article

Keywords: Industrial agglomeration, Spatial autocorrelation, Moran's I, LISA, Principal Component Analysis, Spatial Lag Model

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10764785/v1

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

Spatial Patterns and Determinants of Industrial Operation in Maharashtra: A District-Level Spatial Econometric Analysis

Sudeb Pal[1*] and Sumana Bandyopadhyay[2]

[1]UGC-Senior Research Fellow, Department of Geography, University of Calcutta, Kolkata, West Bengal, India, PIN- 700019
sudebcu@gmail.com

Sudeb Pal: ORCID ID: 0000-0001-9165-3661

[2]Professor at Geography, Department of Geography, University of Calcutta, Kolkata, West Bengal, India, PIN-700019
sumona_bm@yahoo.com

Corresponding Author*:sudebcu@gmail.com

Abstract

The industrial activity in Maharashtra is characterized by high concentration in space and regional disparities but the contribution of the industrial characteristics and the neighbouring districts to such a pattern is not sufficiently understood. This study utilizes the Annual Survey of Industries (ASI) 2023–24 data at the district level to investigate the spatial distribution and characteristics of operational manufacturing industries across the districts of Maharashtra. We use a combination of Principal Component Analysis (PCA), Moran's I, LISA and spatial regression to identify major dimensions of industrial activity and to account for spatial dependence. PCA summarizes the highly correlated industrial indicators into four principal components explaining 95.96% of the total variance. Strong associations were found for physical working capital, profit, employment, gross value added, output and capital related indicators. The spatial analysis has shown significant clustering of operational industries. High–High clusters are located around Palghar, Thane, Pune, Raigad and Mumbai Suburban, whereas Low–Low clusters are seen in a number of districts of Marathwada and Vidarbha. The spatial regression diagnostics suggest a Spatial Lag Model (SLM) rather than the classical OLS specification, suggesting that the industrial operation is related to the industrial activity of neighbouring districts. The chosen SLM also performs much better regarding fit of the model with $R^2 = 0.9963$, log-likelihood = −182.810, AIC = 403.621, BIC = 421.040 and $\sigma^2 = 3{,}103.99$ than the OLS model. The results suggest that industrial geography of Maharashtra is determined by the conjoint effect of industrial capacity, capital, production, employment, agglomeration and spatial dependence. The study emphasizes the importance of spatially differentiated industrial policies to augment secondary industrial centres, connectivity, infrastructure and production linkages in relatively less industrialized districts.

Keywords: Industrial agglomeration; Spatial autocorrelation; Moran's I; LISA; Principal Component Analysis; Spatial Lag Model.

1. Introduction:

Industrialization continues to be one of the most potential driving forces of economic development since it fosters structural changes, employment opportunities, increased productivity, innovation, and competitiveness (Naudé & Szirmai, 2012; Van Neuss, 2019). Manufacturing increases productive links in agriculture and service sectors and provides employment directly and indirectly. Thus, it contributes to economic development that is inclusive. The strategic importance of manufacturing has been recognized and manufacturing-led development has been stressed in the industrial policies of emerging economies that followed as a way to achieve long-term economic growth and reduce regional inequality (Lewis, 1954; Hirschman, 1958; Rodrik, 2013).

Industrial development is important, but geographically varied. Economic activity is expected to be concentrated in areas with adequate transport facilities, skilled labour, institutions and access to local and foreign markets. The spatial concentration leads to the emergence of industrial agglomeration economies which increase productivity and competitiveness but also lead to disparities between the industrially developed and poor regions (Myrdal, 1957; Isard, 1956; Fujita & Thisse, 2002). Therefore, the spatial distribution of industrial activities and the variables such as infrastructure related, labour and capital investment impacting industrial performance have become important issues in regional development planning and industrial policy.

The industrial sector in India has witnessed a phenomenal change since the economic reform and liberalization in 1991 (Kotwal et al. 2011; Bhalotra, 2002). As the pace of industrial expansion has accelerated in different parts of the world, driven by technological advances, market-oriented reforms, higher levels of foreign direct investment and integration into global value chains, the regional disparities in industrial development have also increased. The industrial growth has been concentrated in the states with better infrastructure, institutional capacity, logistics and investment climate as manufacturing has become more capital intensive and technologically advanced (NITI Aayog, 2024; Srivastava et al., 2026). This uneven pattern of industrialization has resulted in serious policy problems of equitable regional development and efficient allocation of industrial resources.

Maharashtra is an important state in the country's industrial sector among Indian states. The state has one of the most diversified industrial bases in the country, including autos, engineering, chemicals, pharmaceuticals, textiles, food processing, petroleum products and

electronics (Ministry of Statistics and Programme Implementation [MoSPI], 2025). It is also an important part of the organized manufacturing sector of India. The Maharashtra Industrial Development Corporation (MIDC) has been instrumental in the uninterrupted industrial growth for several decades through the development of industrial estates, well-developed urban centres, extensive transportation infrastructure, major ports and strategic geographic location (Government of Maharashtra, 2024; MoSPI, 2025). But the industrial growth of Maharashtra is not spatially uniform (Chakraborty, 2024; Kalota & Khan, 2011). While some parts of Vidarbha, Marathwada and northern Maharashtra are still experiencing relatively slow industrial growth, Pune, Thane, Mumbai Suburban, Raigad, Palghar and Nashik have become major manufacturing hubs (ASI report 2023-24). These intra-state variations suggest that the causes of differences in regional industrial performance and industrial operation need to be investigated.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature and builds the conceptual framework. Section 3 describes the study area, data sources and research methodology. Section 4 reports the empirical results and spatial econometric analysis. The fifth section discusses the findings in relation to the existing literature and policy implications and the final section provides the conclusion of the study and directions of future research.

2. Literature Review:

2.1 Regional Development, Agglomeration and Industrial Development:

Industrial development is often seen as important for employment, economic growth, technological progress and regional competitiveness. Marshall (1890) proposed that the advantages of localization economies, labour markets specialized in particular skills, supplier linkages, and knowledge spillovers, lead to the clustering of businesses regionally. These advantages create industrial districts which reduce production costs and increase productivity. The Growth Pole Theory explains that the industrial development begins in the economically active centres and slowly diffuses to the neighbouring regions through backward and forward linkages (Perroux, 1955). Krugman (1991) developed the New Economic Geography which illustrates how increases in market access, transportation costs and returns to scale can lead to a cumulative industrial concentration, often leading to persistent regional inequities. Geographically clustered businesses are more competitive due to innovation, collaboration and shared infrastructure, according to Porter's (1998) cluster theory.

Urbanization and industrial land use planning have increasingly become important determinants of the location of industry and regional competitiveness. Empirical studies have shown that agglomeration in industry is based on labour, raw materials, transport infrastructure, institutional backing, and market proximity (Zhang et al., 2023; Shang, 2026). In addition, capitals are extremely important for the competitiveness of an industry because enough expenditure of funds will result in efficiency in operation, technological development, and growth of industries (Yarova, 2021; Barakat, 2024). Technology innovation and equipment updating are essential for any change and competitiveness of industry (Rüßmann et al., 2015).

2.2 Factors Influencing the Performance and Structural Development of Industry:

The performance of the industry is determined by institutional and economic factors such as labour productivity, capital investment, infrastructure, innovation, access to financing and industrial policy (Oqubay, 2025; Chang & Andreoni, 2020). However, labour and capital productivity continue to be important factors for the expansion of manufacturing. The fast urbanization process has had mixed effects, encouraged the growth of the service sector while reduced the relative importance of manufacturing activities in some areas of the world (Mendez et al., 2023). However, the decisions on the location of the business sites are influenced by manufacturing costs, market access, transportation networks, and consumer demand (Tate et al., 2014). The industrial characteristics of the region are constantly changing due to the growth of the economy, migration, environmental regulations, the industrial structure, and technological development (Ashton, 2009; Wu et al., 2018). The process of globalization drives the structural change of Industry in the industrialized countries. Impacts of deindustrialization, a slow process of changing from industrialized economies to service-oriented economies, have been found on employment, productivity and economic growth (Van Neuss, 2019; Luan, 2025). The growth of the service sector has created new economic opportunities. But many industrial regions have suffered from structural imbalances, unemployment and regional inequality as a result of the demise of the traditional manufacturing industries (Kollmeyer & Pichler, 2013; Kitson & Michie, 2014). Such changes highlight the need to identify what continues to give manufacturing a competitive edge in fast changing regional economies.

2.3 Industrial Analysis: Regional Strategies

The spatial character of industrial development has resulted in the increased use of spatial analytical approaches in regional studies. Many research works have used methods of exploratory spatial data analysis (ESDA) such as Location Quotient, Global Moran's I, Local

Indicators of Spatial Association (LISA), and Getis-Ord Gi* in order to identify industrial clusters, spatial concentration and regional disparities (Carroll et al., 2008; De Siano, 2020). These techniques permit researchers to identify geographical trends and statistically significant clusters of economic activity that are not apparent from traditional descriptive statistics (Wheeler et al., 2013).

Spatial econometric models have become increasingly important in the past years because they consider in a specific way the geographical dependence of adjacent regions. Effects of spillover from the surrounding regions on industrial performance via infrastructure, labour migration, investment, and technology transfer have been investigated in literature using spatial autoregressive (SAR), spatial error model (SEM), and spatial durbin model (SDM) (Anselin, 1988; LeSage & Pace, 2009; Elhorst, 2014; Wang et al., 2023). The field of spatial econometrics becomes extremely crucial for investigating the dynamics of regional economics because empirical evidence proves that the growth of industries seldom occurs in isolation within geographical boundaries (Qu et al., 2023).

The present research tries to bridge this gap by investigating the determinants and characteristics of industrial operations in the districts of Maharashtra by employing district-level data from the Annual Survey of Industries (ASI) 2023–24. In contrast to previous studies, this study employs descriptive statistical analysis, exploratory spatial data analysis, Moran's I cluster analysis, LISA map and Spatial Autoregressive (SAR) modelling to investigate both the determinants of industrial operation and their spatial interdependence. The study integrates industrial characteristics like employment, capital investment, production, value addition, financial performance, and industrial infrastructure into a spatial econometric framework to provide a comprehensive understanding of industrial concentration and regional disparities.

The novelty of this study comes from three major contributions. First, it provides a comprehensive assessment of industrial operation at the district level on the basis of a wide range of industrial indicators rather than a single economic indicator. Second, it explicitly incorporates spatial dependence into the analysis by applying Moran's I and Spatial Autoregressive modelling, thereby addressing a significant methodological limitation of previous research. Third, it determines the influential industrial characteristics for operational manufacturing industries and uncovers the spatial spillover effects that shape regional industrial development. The results are expected to contribute to the literature of industrial

geography and regional development and to provide empirical evidence for formulating balanced industrial policies and for sustainable industrial growth in Maharashtra.

Thus, the objectives of this study are to: study the spatial distribution and regional concentration of operational manufacturing industries in the districts of Maharashtra; determine the dominant industrial characteristics affecting the operation of manufacturing industries at the district level; examine the presence of spatial autocorrelation and industrial clustering using Moran's I and Local Indicators of Spatial Association (LISA) map; and quantify the impact of industrial characteristics on operational manufacturing industries using Spatial Autoregressive (SAR) modelling.

3. Study Area:

The study is conducted in the state of Maharashtra, located in the western peninsular part of India on the border of the Deccan Plateau. It is the third largest in the nation by area and the second most populated. Maharashtra's boundaries are with Madhya Pradesh and Gujarat in the north, Chhattisgarh to the east, Karnataka and Goa in the south, and the Arabian Sea in the west. The state is administratively organized into 36 districts, further divided into six divisions (Figure 1). Mumbai, the state's largest urban agglomeration, is also India's financial hub, and Nagpur is the winter capital. Maharashtra has a very long history of industrialization and is one of India's most industrialized states. Its economy is highly diversified with large-scale industries in manufacturing, engineering, chemicals, pharmaceuticals, automobiles, food processing, and information technology. Most of India's top corporate and financial institutions have their headquarters in Mumbai, reinforcing the state's role as an economic hub. Maharashtra State had a population of more than 112 million according to the 2011 Census, almost one-tenth of the country as a whole. Data published in the Annual Survey of Industries for the year 2023-24 indicates that more than 2.56 million individuals are gainfully employed in registered industries in the state of Maharashtra, out of which around 1.87 million are workers.

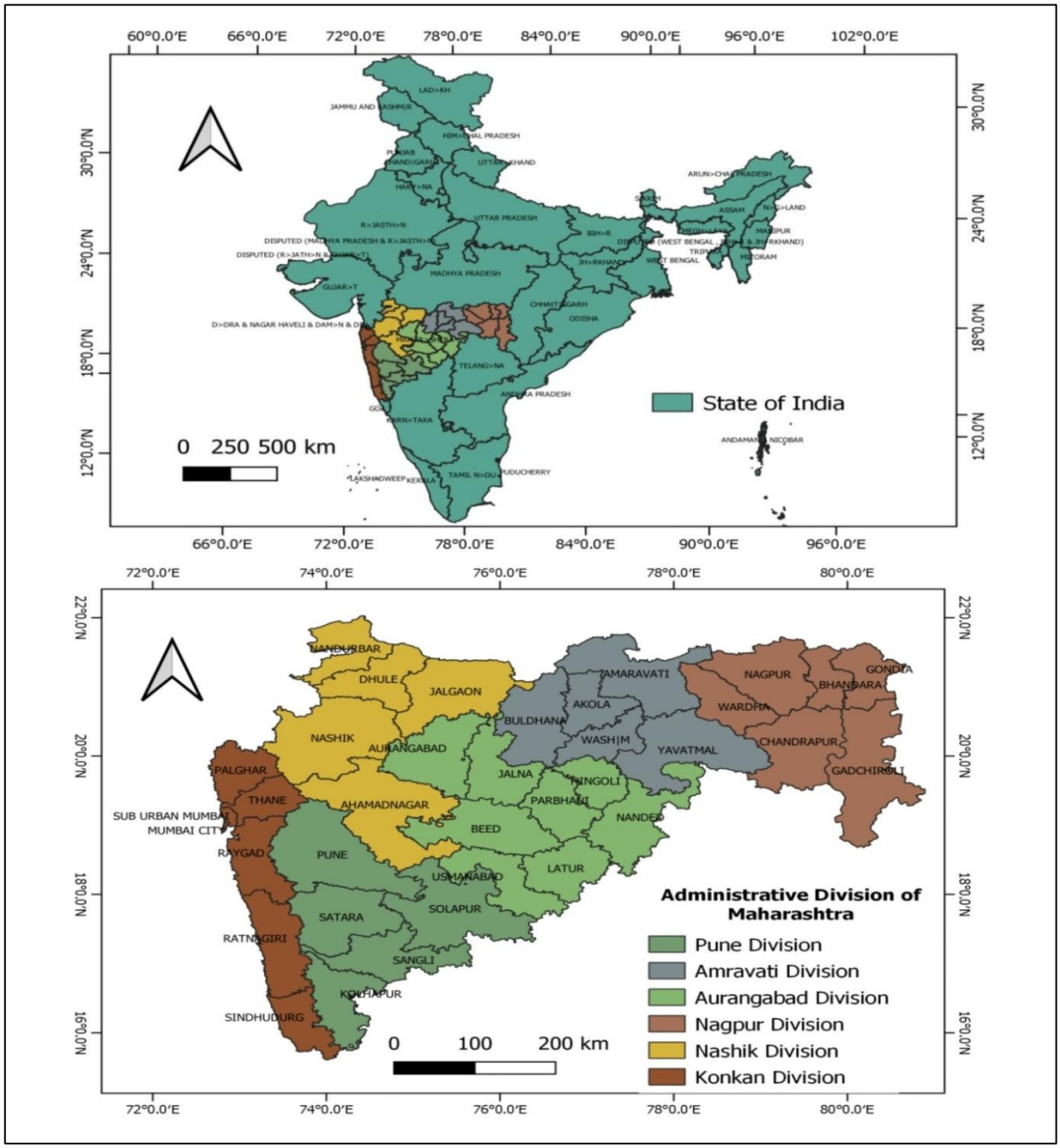


Figure 1: Location Map of the Study Area, **Source:** Survey of India

Maharashtra has been selected as the study area because it offers a combination of industry age, geographic diversity, and economic importance that is unique to it. It never accounts for less than a sizeable proportion of Indian manufacturing output and gross state domestic product. The state covers diverse industrial sectors, ranging from traditional textile to automobiles, pharmaceuticals, and financial services, and discover a wide intra-state variation in industrial development. Industrially developed corridors like the Mumbai-Thane-Pune and Nashik-Aurangabad belts exist in the state, along with comparatively less industrialized areas like Vidarbha and Marathwada (Economic Survey of Maharashtra 2023–24, GoI, 2024). This diversity enables fruitful comparison of the impact of various factors on industrial activity in

various contrasting regional contexts. Its strategic position on the western seaboard, availability of big seaports, high transport connectivity, and huge industrial estates built up by the Maharashtra Industrial Development Corporation (MIDC) further render it a proper context to investigate the operational dynamics of industries. These traits collectively offer a more engaging and representative backdrop within which to analyse the factors and attributes that determine industrial performance.

4. Materials and Methods:

4.1 Data Source and Selection of Variables:

In order to measure the regional and spatial concentration of industries and operational factories out of the total number of registered industries/factories in Maharashtra. This study utilizes data from the Annual Survey of Industries (ASI) for 2023-24, based on pooled data from both central and state samples for Maharashtra (ASI, 2023-24). For instance, 47 variables were used to conduct Principal Component Analysis (PCA) using data from the Annual Survey of Industries. From this industrial data source, we have extracted several explanatory variables, including the number of registered factories, factories in operation, physical working capital, gross value added to fixed capital, gross value of plant and machinery, total output, depreciation, profit, and the number of male workers, which influence the functioning of the industries. Additionally, data on other potential influential factors in operational industries were collected, including the number of workers, employees other than workers, unpaid family members, proprietors, etc., engaged in factories (Table 1); man-days employed in industries; and wages and salaries, including the employer's contribution, were considered in the analysis. GeoDa and SPSS (version 28) software were used for the analysis.

4.2 Data Analysis:

4.2.1 Spatial Autocorrelation Analysis (Moran's I and LISA Map):

The study analyzes the drivers and spatial patterns of functioning industries within the districts of Maharashtra. Spatial autocorrelation, in contrast to regular correlation, is defined as correlation of the same variable based on the geographic (spatial) arrangement of values. It entails spatial clusters, where there is one variable with larger concentrations being surrounded by equal large concentrations, and spatial outliers, where one variable of high value has low concentrations surrounding it. Spatial autocorrelation arises when spatial randomness is broken (Liu et al., 2013; Abdali & Rajaei, 2020). Industrial operations are, however, region specific and tend to cluster in a particular region based on common infrastructure, market access, and

labour pools. Prior to examining the drivers of industrial presence, it is necessary to establish if indeed spatial dependence exists in the data (Banerjee et al., 2025; Anselin, 2022). Moran's I method assesses if the distribution of operational industries is spatially random or spatially grouped, which aids directly the aim of establishing regional concentrations and differences in industrial growth. Local Indicators of Spatial Association (LISA) adds to this by indicating where these clusters exist, identifying high-high, low-low, high-low, and low-high spatial outliers between the 36 districts. Diagnostic for spatial regression can be executed, since omitting spatial dependence would be in violation of the independence of residuals assumption in classic regression and create biased estimators. Moran's I and LISA offer empirical justifications for the utilization of spatial econometric models over the use of common OLS by themselves (Anselin, 1995; Chen, 2024; Majumder et al., 2026).

Moran's I needs a spatial weight matrix to specify the neighborhood around each geographic unit. Each geographical unit's value is compared to the average values of its neighboring units (Qu & Lee, 2015; Anselin et al., 2005). Weights can be calculated in terms of contiguity to the boundary polygon or be derived from interpoint distance. First-order queen contiguity has been used in this research utilizing GeoDa software to the boundary polygons (shape files) in order to calculate the spatial autocorrelation of each geographical unit. This method determines the tendency of similar properties to cluster in certain geospatial sites (Getis & Aldstadt, 2004). The Moran's I index formula is as follows:

$$I_{local} = \frac{n(Xi-\bar{X})\sum_{j=1}^{n} Wij\,(Xj-\bar{X})}{\sum_{i=1}^{n}(Xi-\bar{x})2}$$

[Where n is the total number of spatial units in the sample of the study area, X_i and X_j represent the location of operational industries/ factories in areas i and j, $\bar{x}$ is the average number of industries across 36 districts and Wij is the spatial weight matrix. A positive value of Moran's I indicates that region i is spatially adjacent to regions with similar values. Conversely, a negative value indicates that region i is surrounded by regions with dissimilar values (O'Leary et al., 2016; Jesri et al., 2021). The values of Moran's I range from -1 to 1. A value less than 0 indicates negative spatial correlation among the spatial units, a value greater than 0 indicates positive correlation, and a value of 0 indicates no spatial relationship.]

4.2.2 Principal Component Analysis (PCA):

The industrial dataset was tested for its appropriateness of extracting the principal component using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's Test of Sphericity. The KMO

value of 0.652 is acceptable for sampling adequacy which shows that the industrial indicators share enough variance to perform factor extraction (Constantin, 2014). Bartlett's Test of Sphericity was significant ($\chi 2 = 1461.263$, df = 1081, $p < 0.001$), indicating that the correlation matrix is not an identity matrix and that PCA is appropriate.

PCA was used as an exploratory data reduction method to reveal the latent structure of 47 industrial indicators. Instead of using all of the highly correlated variables in the regression analysis at the same time, the results derived from PCA were used to select representative variables from each principal component. Pearson correlation and Variance Inflation Factor (VIF) diagnostics were then used to test these candidate variables, before estimating the spatial lag model. Thus, PCA was used to reduce redundancy and improve model specification, and not used directly as a predictor in the regression model (Wang et al., 2024; Sarwar et al., 2026).

Table 1: KMO and Bartlett's Test of Sphericity

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure...		0.652
Bartlett's Test of Sphericity	Approx. Chi-Square	1461.263
	df	1081
	Sig.Bartlett	.000

4.2.3 Spatial Auto-regression Models:

Having established spatial dependence, it is necessary to apply models that are capable of handling it. Spatial clustering, agglomeration, and industries' diffusion have marked spatial variability across industrial growth in various regions. Therefore, geographical weighted matrix-based spatial auto-regressive model can strongly analyze the impact of spatial effects on industrial growth and its determinants (Chen et al, 2018). The spatial auto-regression model extends the general linear regression model where the spatial properties of data and spatial weight matrix values are brought into the analytical model. The general form of spatial regression model is as follows (Hua et al, 2021):

$$Y = pW1 + \beta X + \mu \ldots\ldots\ldots\ldots\ldots\ldots \text{(eq. I)}$$

$$\text{where, } \mu = \lambda W2 + \varepsilon$$

[Where, Y is the dependent variable, ρ is the coefficient of spatial non-linearity term, X is the explanatory variable, β explains the spatial regression coefficient of the explanatory variables (X), μ is the error term, ε represents any noise, W_1 and W_2 are geographical weight matrices which reflect the spatial trend and pattern of the dependent variable and residuals, λ is the coefficient of the spatial error term.]

For this research, the functioning of industries/factories have been chosen as the dependent variable, while a group of explanation variables consisting of physical working capital, overall number of registered factories, gross value added to fixed capital, overall output, depreciation, profit, and overall number of male workers have been employed as independent variables. These variables have been tested with the spatial regression model by applying the spatial weighting technique outlined above.

4.2.3.1 Ordinary Least Square Method (OLS):

The Ordinary Least Square (OLS) model is a common technique of linear regression analysis that is utilized for detecting a linear relationship between an explanatory (independent) variable set and a dependent variable (Craven & Islam, 2011; Dismuke & Lindrooth, 2006). OLS is efficient statistically if there is no spatial dependence. It is the diagnostic standard to check whether more advanced spatial models are required. This approach is identified as the Best Linear Unbiased Estimator [BLUE] (Puntanen & Styan, 2011). The OLS model holds when ρ = 0 and λ = 0, which means that the model remains unaffected by geographical distribution (Hua et al., 2021). The OLS estimates the spatial analysis coefficients by minimizing the sum of squared prediction errors, and for this reason, it is called "least squares." The equation is as follows:

$$y = X\beta + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots \text{(eq. II)}$$

4.2.3.2 Spatial Lag Model (SLM):

The Spatial Lag Model (SLM) of spatial statistics and econometrics maintains spatial contiguity in varied order of magnitudes (Anselin & Smirnov, 1996). Thus, a dependent variable y at place i is greatly influenced by the independent variables at places i and j. This model is developed on the basis of spatial autocorrelation where ρ≠0, λ= 0 (Hua et al, 2021). The primary aim of building Spatial Lag Model (SLM) is to measure the spillover impacts of the chosen variables on the neighboring geographical locations (Golgher & Voss, 2016). Using

a spatially lagged dependent variable, SLM addresses neighborhood effects and spatial spillovers, automatically in accordance with the aim of examining spatial clustering and regional diffusion of industries. The formula applied in this model is as follows:

$$Y = \rho Wy + X\beta + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots \text{ (eq. III)}$$

[Where *W*=spatial weight matrix, *Wy* =spatial lag dependent variable, ρ= spatial regression coefficient and β= regression coefficient in the model.]

4.2.3.3 Spatial Error Model (SEM):

When the spatial regression coefficient is zero and the coefficient of the spatial error term is unequal to zero, then a spatial error model can be derived. To prevent the spatially correlated errors induced by omitted variables in the spatial data set (Golgher & Voss, 2016; Baltagi et al., 2003). It has been used whenever unobserved spatially correlated factors (such as state policies, infrastructure quality, or institutional capacity) affect industrial location, and their exclusion from the model induces spatially correlated residuals. The model has utilized the following formula as follows:

$$Y = X\beta + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots \text{ (eq. IV)}$$

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

[Where λ is the spatial error coefficient that explains the degree of influence between the region i and j, ε represents the error term with spatial autocorrelation, and μ is the polygon expressing the independency of the normal distribution.]

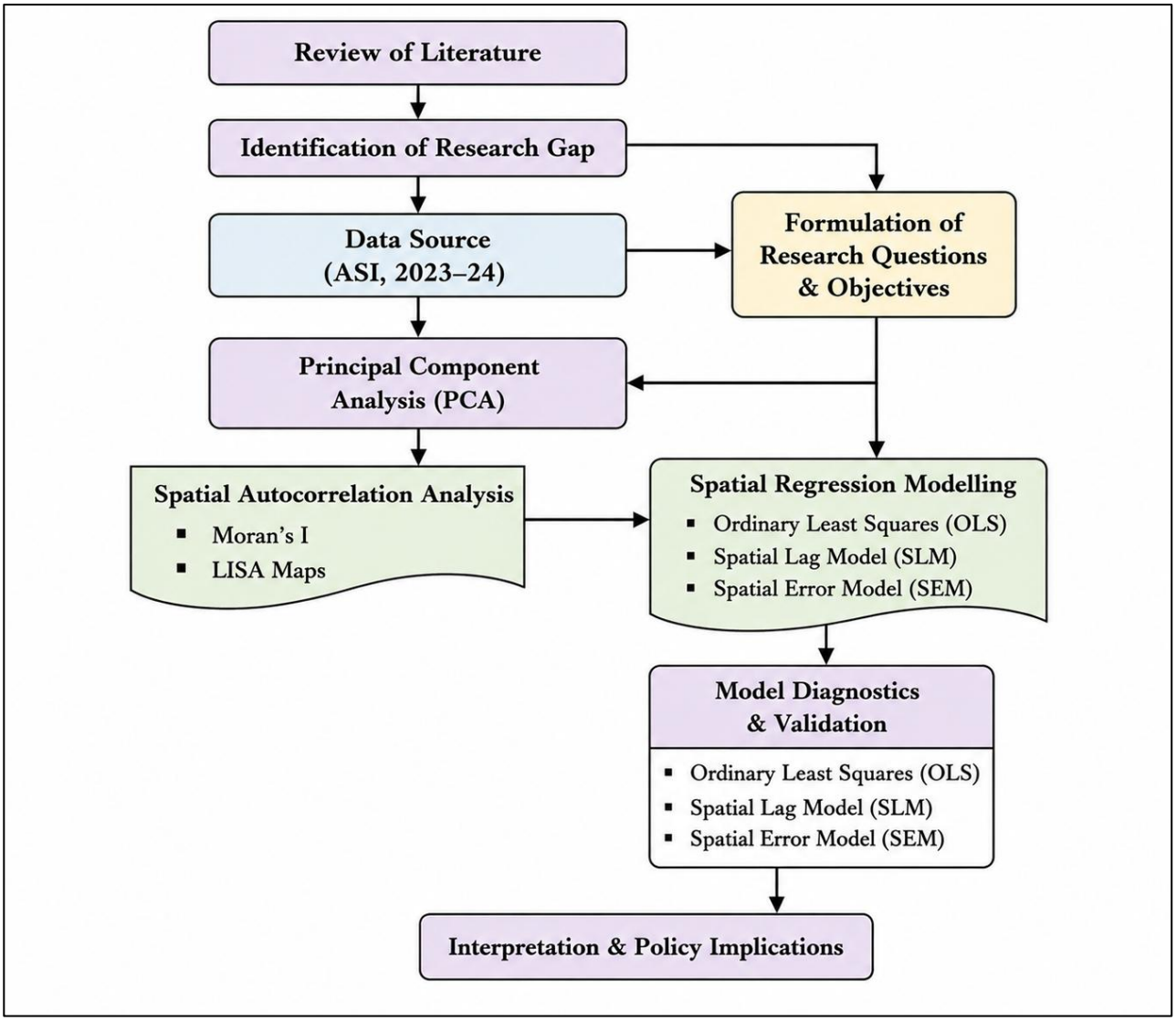


Figure 2: Methodological Flowchart (Prepared by Author)

5. Results:

5.1 Descriptive Statistics of Employment Characteristics:

The employment characteristics of the sample population are presented in (Table 1). Descriptive statistics show that the mean number of persons engaged in industries/factories such as manufacture of fabricated metal products, food products, machinery, chemicals and chemical products, equipments, and manufacture of textiles in Maharashtra are 59080.03 (SD= 112786.605). Out of the total number of persons involved in industries, workers were 72.97 per cent, employees other than workers were 26.91 percent, unpaid family members/proprietor etc. were engaged about 0.12 per cent respectively. Workers can be employed either directly or through the contractors.

Table 2: Descriptive Statistics of Employment Characteristics

Characteristics	Range	Minimum	Maximum	Mean	Std. Deviation
A. No. of Persons Engaged	787090	684	787774	71074.69	143838.158
1. Workers (no.)	575900	475	576375	51863.72	104342.718
1.1 Directly employed (no.)	213428	380	213808	24710.81	41896.182
1.1.1 Men (no.)	189072	343	189415	21990.42	37167.858
1.1.2 Women (no.)	24392	0	24392	2720.31	5231.083
1.2 Employed through Contractors(no.)	362524	43	362567	27152.97	63566.222
2. Employees other than workers (no.)	210956	194	211150	19128.97	39614.888
2.1 Supervisory & Managerial Staff (no.)	86961	48	87009	8023.36	16175.269
2.2 Other Employees (no.)	124039	102	124141	11105.67	23523.926
3. Unpaid Family Members / Proprietor etc. (no.)	567	0	567	82.17	124.249

Data Source: Computed by author from Annual Survey of Industries (ASI, 2023-2024)

These workers may be employed with or without any knowledge of any principal employers or owners and head of the any companies and industries. In this case, mean number of directly employed persons were lower in contrast to mean of employed through contractors (Mean 27152.97 and SD= 63566.222). It is noticeable that men workers are comparatively higher than women workers in the manufacturing industrial organization and evidence from a few numbers of industries showed that the industry was not engaged in a single number of women workers (Minimum=0) where minimum number of men worker was 343. Possible reason for this may include handling heavy machinery, to carry raw materials and by products physically, and for the safety and security purposes. The mean number of employees other than workers is 19128.97 (SD=39614.888). This is inclusive of persons holding position of supervisory, management, engaged in administrative office, store keeping section, welfare section, security

watch and ward staff, sales department, and those engaged in the purchase of raw materials in the factory (ASI, 2023-24). Mean of supervisory and managerial staff is 8023.36 (SD= 16175.269) and mean of other employees are 11105.67 (SD=23523.926) which is comparatively higher than the value of managerial staff. A small number of persons include in the category of unpaid family members and proprietors (Mean=82.17, SD= 124.249) who actively engaged in the work of the factory even without any pay and unpaid members of the co-operative societies who work directly or any productive capacity. Persons in the head office involved in the manufacturing process of the industries are also counted in this category.

5.2 Principal Component Analysis:

Four principal components were extracted by Kaiser criterion (Eigenvalue > 1) that explained 95.96% of the total variance. The first component alone explained 83.21% of the variance, suggesting that most of the industrial indicators are tapping into a common latent dimension that captures the general level of industrial development. The component matrix was rotated by Varimax method to improve the interpretability. Only factor loadings ≥0.60 are presented. The rotated solution reveals four dimensions of industrial operation.

Table 3: Total Variance Explained

Component	Eigenvalue	Variance Explained (%)	Cumulative (%)
PC1	39.107	83.206	83.206
PC2	3.500	7.447	90.653
PC3	1.460	3.105	93.759
PC4	1.035	2.202	95.961

Table 4: Rotated Principal Component Matrix (Loadings ≥ 0.60)

Industrial Variables	PC1	PC2	PC3	PC4
Physical Working Capital	0.930			
Profit	0.922			
Men Workers	0.907			
Number of Factories	0.892	0.705		
Depreciation	0.891			

Industrial Variables	PC1	PC2	PC3	PC4
Gross Value Added	0.879			
Net Income	0.878			
Wages & Salaries	0.877			
Women Workers	0.874			
Wages including Bonus	0.876			
Interest Received	0.859			
Supervisory Staff	0.886			
Directly Employed	0.805			
Factories in Operation	0.835			
Employer Contribution	0.828			
Employed through Contractors	0.821			
Employees other than Workers	0.797			
Value of Product & By-products	0.792			
Total Output	0.756	0.887		
Fuels Consumed		0.862		
Gross Fixed Capital Formation		0.781		
Gross Capital Formation		0.776		
Gross Value Addition to Fixed Capital	0.601	0.770		0.795
Fixed Capital		0.755		
Invested Capital	0.700	0.701		
Plant & Machinery	0.685	0.691		
Materials, Fuels etc. (Stock)			0.837	
Unpaid Family Members			0.823	
Addition in Stock			0.752	
Finished Goods				−0.917

Extraction Method: Principal Component Analysis (PCA); Rotation Method: Varimax with Kaiser Normalization

The first principal component (PC1) has high loadings on physical working capital, number of factories, factories in operation, employment, wages, gross value added, profit, depreciation, and net income. These variables capture the operational scale, employment intensity, financial performance and productive capacity of manufacturing industries. PC1 is thus interpreted as the dimension industrial operation and economic performance. The second main component (PC2) is mainly characterized by the total output, fuel consumption, fixed capital, invested capital, plant and machinery, and gross capital formation. These variables are typically measures of production capacity and long-term capital investment, which are indicative of areas with concentrated manufacturing activity and significant industrial investment. PC2 is therefore the production and capital investment dimension.

Stock of materials, unpaid family members and addition in stock load high positively on third principal component (PC3). These variables are particularly indicative of the inventory management and ownership patterns in relatively smaller industrial establishments. Therefore, PC3 is interpreted as inventory and small-scale ownership characteristics.
The fourth principal component (PC4) is largely associated with gross value addition to fixed capital and finished goods where the latter has a strong negative loading. This component captures the trade-off between the expansion of capital and the accumulation of inventory and is interpreted as the capital expansion dimension.

Briefly, the PCA reveals that four latent dimensions: operational efficiency, capital investment, inventory management and capital expansion are driving the industrial activity of Maharashtra. The strong clustering of employment, capital, production and profitability variables in the first component is also indicative of the high degree of interdependence of industrial characteristics across districts.

In contrast, the scree plot shows the distribution of eigenvalues for the 47 principal components extracted. We find a steep decrease of eigenvalues from the first to the second component and a slow flattening of the curve, which indicates the diminishing marginal contributions of the later components. The first principal component captures the primary source of variation among the industrial indicators, and the remaining components account for relatively smaller shares of the variance. According to the Kaiser criterion (eigenvalue > 1), four main components were retained for further interpretation. The four components together explain 95.96 % of the total variance. The first component accounts for 83.21 % of the variance. The sharp elbow at the fourth component indicates the components after the fourth component add

little additional information, and thus were not used for further analysis.

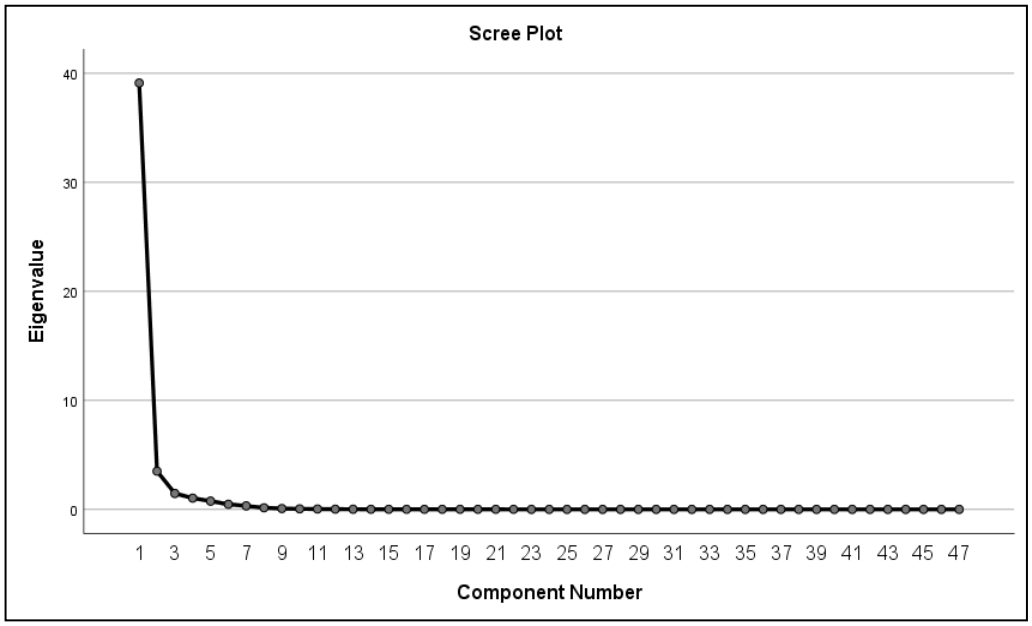


Figure 3: Scree Plot

These components reflect different facets of industrial development, such as operational performance, production and capital investment, inventory management and capital expansion. Since the number of industrial indicators (47 variables) is relatively large compared to the sample size (36 districts), therefore, the PCA is interpreted as an exploratory data reduction technique rather than a confirmatory factor model. The extracted components were utilized to identify clusters of highly correlated industrial features and to inform the choice of representative variables for subsequent spatial regression analysis.

5.3 Spatial Autocorrelation Analysis using Moran's I Values:

Before estimating the spatial regression model, global Moran's I statistics were calculated to verify the spatial dependence of industrial activities across the districts of Maharashtra. The analysis used a first-order Queen contiguity spatial weights matrix, which assumes that districts which are neighbours (i.e. share a common boundary or common vertex) influence each other. Testing for spatial autocorrelation is important as the classical Ordinary Least Squares (OLS) regression assumes independent observations. If there is significant spatial dependence, this assumption is violated, and spatial econometric models are justified. Moran's I Operating Factories (Figure 4) shows three Moran's I scatter plots for Factories in Operation (Var2) in different spatial situations. Figure 4A provides data without major industrial concentration districts, yielding a Moran's I of 0.091, suggesting only weak positive spatial autocorrelation.

The almost horizontal regression line suggests that the spatial dependence is quite limited when the dominant industrial districts are absent. The complete district dataset is shown in Figure 4B, with a Moran's I of 0.325, indicating a moderate positive spatial autocorrelation. The positive slope of the regression line shows that districts with higher numbers of operational factories tend to be surrounded by districts with similarly high industrial activity, while districts with lower numbers of operational factories tend to be surrounded by districts with similarly low industrial activity. The observations are concentrated near the origin, indicating that most districts have moderate industrial activity, while a few districts form high-value clusters. Effect of only major Industrial districts is shown in Figure 4C which resulted in a negative value of the Moran's I (−0.235). This result indicates that the spatial relationship is significantly changed by removing the remaining districts, indicating that the highly industrialized districts are spatial outliers relative to the rest of Maharashtra rather than an independent homogeneous cluster. The comparison of the three scatter plots evidently shows that the observed positive spatial autocorrelation in Maharashtra is contributed mainly by a small number of highly industrialized districts. Thus, industrial concentration is not random or evenly distributed across the state, but is dominated by established industrial growth centres.

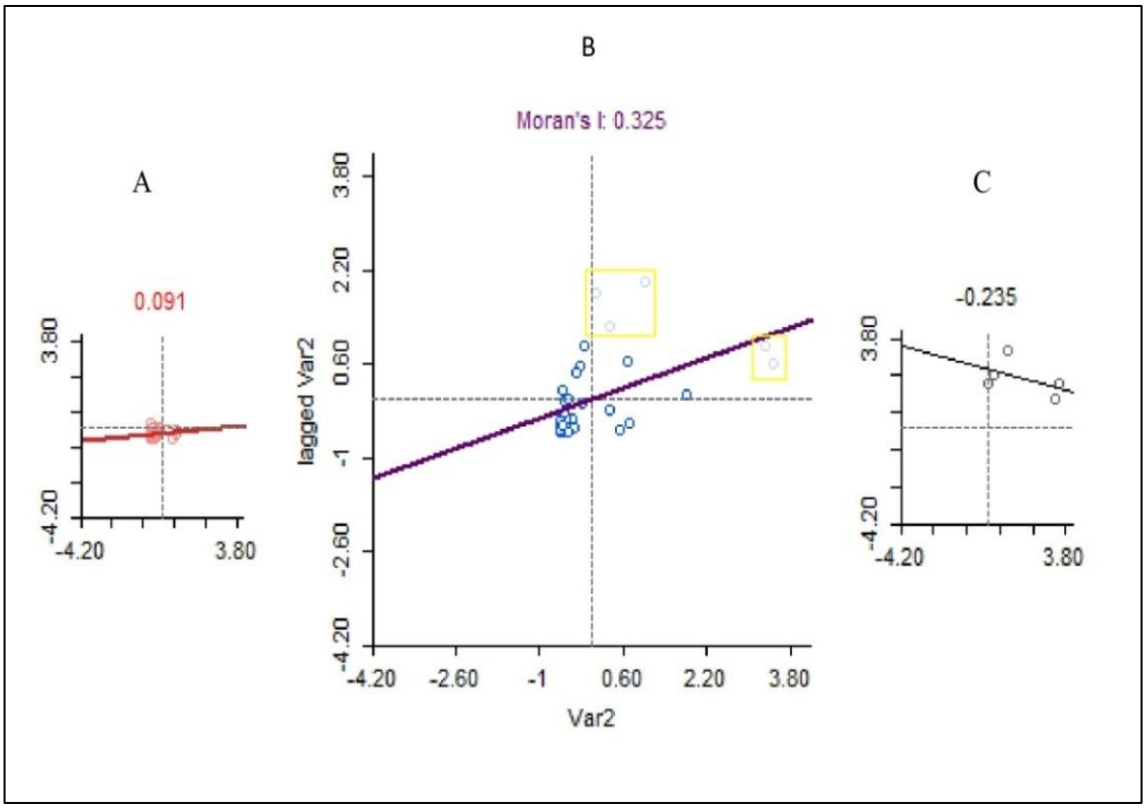


Figure 4: Moran's I Scatter Plot of Industries and Factories in Operation (Variable 2)

5.4 Spatial Dependence of Selected Industrial Characteristics:

All cases show positive slopes in Moran scatter plots, suggesting that all selected industrial characteristics have positive spatial dependence. But the degree of clustering varies across variables. The maximum value of the Moran's I is 0.326 for the number of registered factories and 0.325 for factories in operation indicating that industrial establishments show the highest degree of geographical concentration. This finding is consistent with the presence of industrial corridors and agglomeration economies and indicates that industrial sites are affected by neighbouring districts.

Similarly, there is moderate spatial clustering in physical working capital (0.305) and male industrial employment (0.292), indicating that capital investment and industrial labour are concentrated in neighbouring districts. Such a pattern reflects the tendency of industries to locate where financial resources, skilled labour and supporting infrastructure already exist. The spatial dependence is comparatively weak in case of production related indicators like gross value addition to fixed capital (0.280), depreciation (0.255) and total output (0.240). Production activities are still clustered geographically but seem to be affected not only by spatial proximity but also by firm-specific technological and operational characteristics. Profit has the lowest Moran's I (0.197) among all variables, suggesting that profitability is much less spatially clustered than industrial establishments or capital investment. This implies that the internal efficiency of firms, the structure of their production, market conditions and management practices are more important for the financial performance of firms than the performance of nearby industrial sectors.

The higher values of Moran's I for registered factories, operational factories, physical working capital and industrial employment suggest that these variables are mainly influenced by location-specific factors such as industrial infrastructure, transportation networks, supplier linkages, labour availability and market accessibility. Therefore, these variables have stronger agglomeration effect. In contrast, the lower values of Moran's I of output, depreciation and profit imply that the financial and production performance are less spatially dependent and also affected by firm level characteristics including technology adoption, productivity, managerial efficiency and market competitiveness. This difference indicates that industrial location decisions are more spatially dependent than industrial performance outcomes, which suggests distinct mechanisms regulating industrial concentration and economic performance.

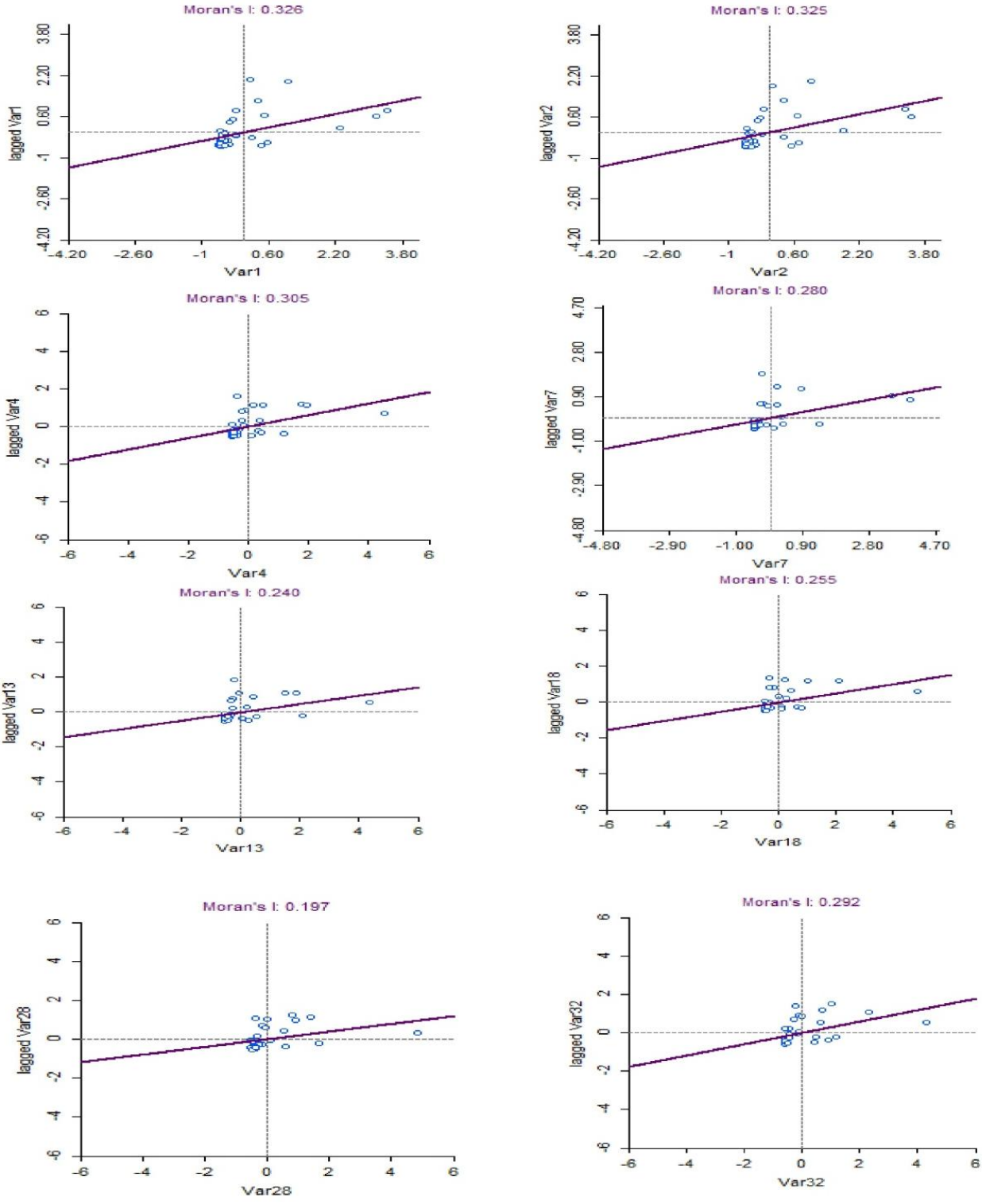


Figure 5: Presents Moran's I scatter plots for eight industrial variables (ASI, 2023-24)

Note: Number of registered industries/factories (Var1), Industries and factories in operation (Var2), Physical working capital (Var4), Gross value added to fixed capital (Var7), Total output (var13), Depreciation (Var18), Profit (Var28), Men number of workers (Var32).

The statistically positive values of Moran's I for all selected variables show that industrial characteristics are spatially correlated and not independently distributed. The assumption of independent observations, as required for Ordinary Least Squares regression, is thus violated. These results thus provide empirical evidence for the estimation of the Spatial Lag Model (SLM) that explicitly takes into account spatial spillover effects between neighbouring districts. Considering spatial dependence helps in more robust estimation of determinants affecting operational manufacturing industries in Maharashtra.

5.5 Local Indicators of Spatial Association (LISA) Cluster Analysis of Industries in Operation:

The LISA (Local Indicators of Spatial Association) cluster map shows the local spatial dependence by depicting the clusters of operational manufacturing industries in Maharashtra, which are statistically significant (Figure 6A). Moran 's I confirms the presence of an overall positive spatial autocorrelation (Moran's I = 0.325). The LISA map reveals the districts that drive this clustering pattern and differentiates between areas with similar and contrasting industrial characteristics.

The analysis indicates five High–High (HH) districts including Palghar, Thane, Pune, Raigad and Mumbai Suburban. Thus, districts with high number of operational industries are surrounded by neighbouring districts with high industrial activity. These districts are the main manufacturing hub of Maharashtra and the most industrialized corridor of the state running from the Mumbai Metropolitan Region towards Pune. These industries have been spatially concentrated due to cumulative agglomeration processes that have been facilitated by improved transport infrastructure, industrial estates set up by the Maharashtra Industrial Development Corporation (MIDC), proximity to ports, large labour markets and diversified manufacturing activities. The clustering of high-value neighbouring districts indicates considerable spatial spillover effects, in which the growth of industry in one district encourages the growth of industry in neighbouring districts.

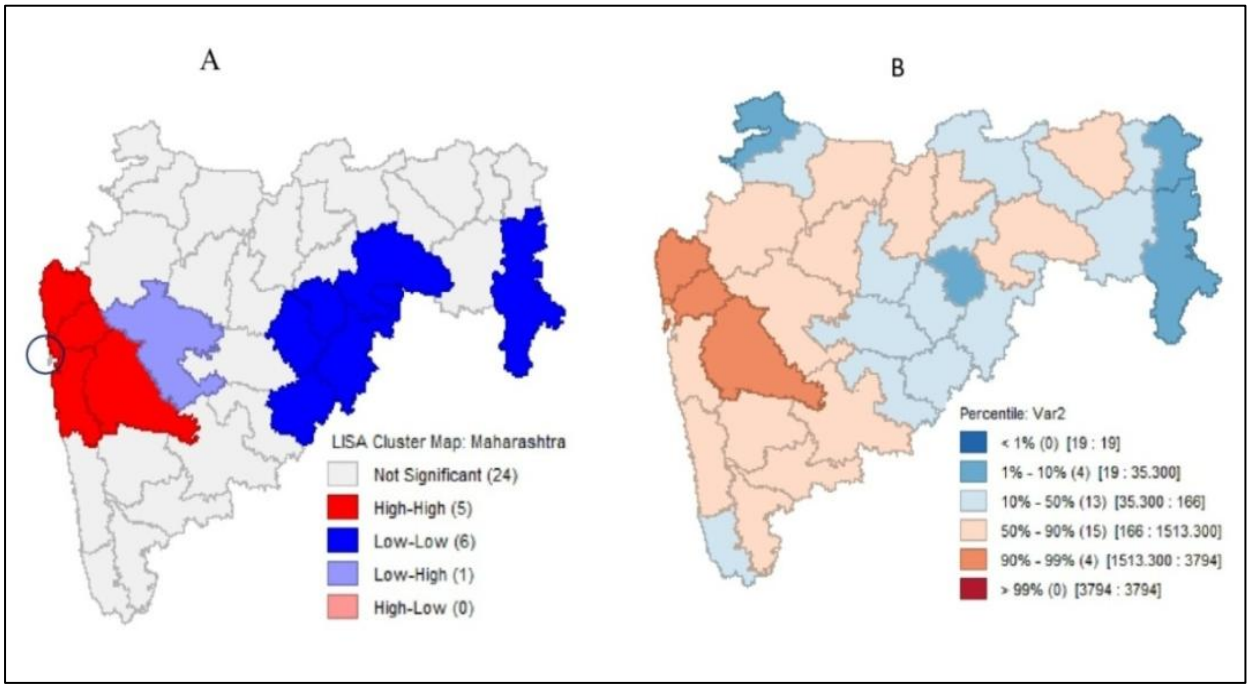


Figure 6: LISA Cluster Map of Industrial Operations (var2)

In contrast, six districts i.e. Hingoli, Parbhani, Nanded, Latur, Yavatmal and Gadchiroli are statistically significant Low-Low (LL) clusters, which are contiguous regions with relatively low levels of industrial operation surrounded by less industrialized neighbours. These districts are concentrated in the eastern and south-central parts of Maharashtra, i.e. Vidarbha and the Marathwada regions. Low-Low clusters are still present, which reflects the spatially uneven diffusion of industrial development and the fact that adjacent districts face similar structural constraints, such as low industrial investment, weak manufacturing bases and limited industrial agglomeration. Also, the analysis identifies one Low-High (LH) district is Ahmednagar (Ahilyanagar). This means that the number of operational industries in Ahmednagar is comparatively lesser but there are many neighbouring districts that have much higher industrial activity. Such spatial outliers are common in transition zones between highly industrialized and less industrialized regions and districts that could potentially benefit from regional spillover effects if connectivity and industrial linkages are improved.

Importantly, no statistically significant HL clusters were detected. Because there are no High-Low districts, heavily industrialized districts are generally absent from low-industrial neighborhoods. The pattern of industrial activity in Maharashtra however is a contiguous pattern of spatial concentration which lends support to the evidence of regional industrial agglomeration rather than isolated industrial centres. This is further supported by the percentile

distribution of operational industries in districts (Figure 6B) and the LISA analysis. The percentile map indicates the relative strength of industrial activity, but it does not consider the spatial relationships between neighboring districts. Therefore, clusters of districts grouped statistically can have different industrial values. The LISA map overcomes this limitation by considering the spatial dependence and statistical significance of industrial concentrations and not their quantitative size.

The overall results of LISA reveal that industrial structure of Maharashtra is characterized by high spatial concentration rather than balanced regional diffusion. Industrial development remains focused on the western metropolitan corridor with continued industrial underdevelopment in the eastern districts. The existence of High-High and Low-Low clusters indicates that the neighbouring districts share similar industrial characteristics, confirming the existence of positive local spatial autocorrelation. These results are consistent with the global Moran's I results, and provide empirical support for the use of spatial econometric models in the subsequent analyses, as traditional regression methods would ignore the spatial dependence.

6. Spatial Regression Analysis (SRA):

6.1 Ordinary Least Squares (OLS) Model and Diagnostics Tests:

First, Ordinary Least Square (OLS) regression was estimated to find out the industrial characteristics associated with spatial distribution of operational manufacturing industries over the districts of Maharashtra. The dependent variable is the number of operating factories (Var2). The explanatory variables are selected according to PCA results and industrial relevance. Table 5 shows the result that registered factories (Var1), gross value added to fixed capital (Var7), profit (Var28) and male workers (Var32) have positive and statistically significant effects on the number of operational industries ($p < 0.05$). In contrast, total output (Var13) and depreciation (Var18) have important negative coefficients and physical working capital (Var4) is statistically insignificant.

Table 5: Ordinary Least Square (OLS) Model

Variable	Coefficient	Std. Error	t-Statistic	*p*
Constant	11.7189	5.72166	2.04816	0.05003
Var1	0.591118	0.0312437	18.9196	0.00000
Var4	0.000123223	8.49075e-05	1.45126	0.15782
Var7	0.000141664	5.56027e-05	2.54779	0.01662

Var13	-2.14942e-05	7.58119e-06	-2.83521	0.00841
Var18	-0.000966984	0.000449041	-2.15344	0.04004
Var28	0.000269485	6.46913e-05	4.16571	0.00027
Var32	0.00922965	0.00141746	6.51138	0.00000

Data Source: Annual Survey of Industries (ASI, 2023-24)

Note: Number of registered industries/factories (Var1), Physical working capital (Var4), Gross value added to fixed capital (Var7), Total output (var13), Depreciation (Var18), Profit (Var28), Men number of workers (Var32).

6.2 Spatial Autocorrelation Tests:

After the OLS estimation, a set of diagnostic tests was performed in an effort to detect the presence and type of spatial dependence (Table 6). The residual Moran's I statistic is negative but not statistically significant (I = –0.0805; p = 0.796), indicating that the residuals from the OLS model are not significantly spatially autocorrelated. However, Moran's I alone is not sufficient to select a suitable spatial econometric model.

Table 6: Spatial Dependence Test Results

Test	MI/DF	Value	Probability
Moran's I (error)	-0.0805	-0.2590	0.79560
Lagrange Multiplier (lag)	1	8.2898	0.00399
Robust LM (lag)	1	8.1415	0.00433
Lagrange Multiplier (error)	1	0.4420	0.50615
Robust LM (error)	1	0.2937	0.58784
Lagrange Multiplier (SARMA)	2	8.5835	0.01368

The Lagrange Multiplier (LM) diagnostics are more persuasive. The LM-Lag statistic (8.2898; p = 0.00399) and Robust LM-Lag statistic (8.1415; p = 0.00433) are statistically significant, whereas the LM-Error and Robust LM-Error tests are not (p > 0.50). Moreover, the SARMA statistic (8.5835; p = 0.01368) corroborates the presence of spatial dependence. According to the model selection framework proposed by Anselin (2005), the Spatial Lag Model (SLM) should be selected instead of the Spatial Error Model (SEM) if both the simple and robust LM-Lag statistics are significant, and the corresponding LM-Error statistics are insignificant. We therefore have reason to estimate a spatial lag specification based on these diagnostic results.

6.3 Spatial Lag Model (SLM):

In Spatial Lag Model the impact of spatial interaction is explicitly modelled by using a spatially lagged dependent variable. The estimated spatial autoregressive coefficient (ρ) is negative and

statistically significant (ρ = –0.028; p = 0.001) suggesting that there is a significant influence from neighbouring districts on the number of operational industries. Unlike the positive spatial autocorrelation in the Moran's I analysis, the negative spatial lag coefficient should not be taken at face value as an indication of competition between adjacent districts. Rather the implication is that, after controlling for registered factories, capital formation, profitability, employment and other industrial characteristics, districts with relatively higher industrial concentration tend to be associated with a comparatively lower level of industrial operation in adjacent districts. This is capturing the redistribution of industrial activity net of the observed explanatory variables rather than overall clustering pattern. This is consistent with the concentration of manufacturing activity in dominant industrial centres, where investment, infrastructure, skilled labour and industrial services continue to be concentrated in a limited number of districts. Therefore, neighbouring districts might have relatively weaker industrial expansion after controlling for these structural characteristics. Among the explanatory variables, registered factories (Var1) remain the strongest determinant of operational industries, followed by gross value added to fixed capital (Var7), profit (Var28) and male workers (Var32), all with positive and statistically significant coefficients. Physical working capital is still not statistically significant. Total Output still has a negative association with operational industries. Depreciation is only slightly important. The results indicate that the functioning of industry in Maharashtra state is more dependent on the presence of industrial infrastructure, profitability and employment, than on the current level of production.

Table 7: Spatial Lag Model (SLM)

Variable	Coefficient	Std. Error	z-value	Probability
W_Var2	-0.0283513	0.00862798	-3.28598	0.00102
Constant	22.6959	5.53714	4.09885	0.00004
Var1	0.626038	0.0263767	23.7345	0.00000
Var4	3.03127e-05	7.13569e-05	0.424803	0.67098
Var7	0.000230806	5.08132e-05	4.54224	0.00001
Var13	-3.22382e-05	6.72074e-06	-4.79683	0.00000
Var18	-0.000663692	0.000359204	-1.84767	0.06465
Var28	0.000250945	5.03351e-05	4.98548	0.00000
Var32	0.0102451	0.00113802	9.00257	0.00000

Data Source: Annual Survey of Industries (ASI, 2023-24)

Note: Number of registered industries/factories (Var1), Physical working capital (Var4), Gross value added to fixed capital (Var7), Total output (var13), Depreciation (Var18), Profit (Var28), Men number of workers (Var32).

6.4 Spatial Error Model (SEM):

The Spatial Error Model assumes spatial dependence in the error term, not in direct interaction between the neighbouring districts. As shown in Table 8, the spatial error coefficient ($\lambda = -0.438$) is marginally significant ($p = 0.061$) and the signs and significance of the explanatory variables are similar to those estimated in the SLM. The empirical evidence for the SEM is relatively weaker than for the Spatial Lag Model, as the earlier LM diagnostics did not support the presence of spatial error dependence. The relatively low significance of LAMBDA suggests that unobserved spatially correlated factors account for less of the industrial variation than direct spatial interaction between neighbouring districts.

Table 8: Spatial Error Model (SEM)

Variable	Coefficient	Std. Error	z-value	*p*
CONSTANT	14.7089	3.64423	4.03621	0.00005
Var1	0.602488	0.0262985	22.9096	0.00000
Var4	6.44078e-05	6.84437e-05	0.941033	0.34669
Var7	0.000142475	4.79091e-05	2.97386	0.00294
Var13	-2.32832e-05	6.42619e-06	-3.62318	0.00029
Var18	-0.00071755	0.000389233	-1.8435	0.06526
Var28	0.000251697	5.52563e-05	4.55508	0.00001
Var32	0.00952058	0.00121399	7.84238	0.00000
LAMBDA	-0.437603	0.233745	-1.87214	0.06119

Data Source: Annual Survey of Industries (ASI, 2023-24)

Note: Number of registered industries/factories (Var1), Physical working capital (Var4), Gross value added to fixed capital (Var7), Total output (var13), Depreciation (Var18), Profit (Var28), Men number of workers (Var32).

6.5 Comparing and Choosing Models:

The spatial diagnostic test results uniformly suggest the Spatial Lag Model is the best specification to explain the industrial operation in Maharashtra. The choice of the SLM is supported by three complementary pieces of evidence: 1) significant LM-Lag and Robust LM-Lag statistics; 2) LM-Error and Robust LM-Error statistics with negligible; 3) significant spatial autoregressive coefficient(ρ). The results show that the industrial operation is affected not only by the industrial characteristics at the district level, but also by the neighbouring districts via spatial interaction.

Table 9. Comparison of OLS, Spatial Lag Model (SLM), and Spatial Error Model (SEM)

Model	R^2	Log-Likelihood	AIC	BIC	$Sigma^2$	Standard Error	Model Assessment
OLS	0.9839	-222.254	464.509	480.344	18,676.30	136.66	Baseline model
Spatial Lag Model (SLM)	0.9963	-182.810	403.621	421.040	3,103.99	51.71	Preferred model
Spatial Error Model (SEM)	0.9963	-195.921	411.842	427.677	3,123.00	55.88	Alternative spatial model

The OLS model explains a high proportion of the variation in operational industries (R^2 = 0.9839). However, both spatial regression models further improve the explanatory power, with the coefficient of determination increasing to 0.9963, which suggests that spatial processes play an important role in explaining regional variations in industrial operations. Such improvement indicates that neighbouring districts are not independent and that industrial activity is subject to spatial interactions. The fit of the model statistics strongly supports the use of the Spatial Lag Model. The log-likelihood increases substantially from −222.254 in the OLS model to −182.810 in the SLM, indicating a considerable improvement in the goodness-of-fit. The SEM is also an improvement over the OLS model (Log-Likelihood = −195.921), yet it does not fit the data as well as the SLM. Likewise, Akaike Information Criterion (AIC) reduces from 464.509 in OLS model to 403.621 in SLM, and Bayesian Information Criterion (BIC) falls from 480.344 to 421.040. Both values are lower than those found for the SEM (AIC = 411.842; BIC = 427.677), confirming that the SLM provides the best balance between fit of the model and complexity.

The residual variance and prediction error statistics provide further evidence. The residual variance ($Sigma^2$) reduces dramatically from 18676.30 in the OLS model to 3103.99 in the SLM, and the standard error of regression reduces from 136.66 to 51.71. The SEM also reduces the residual variance ($Sigma^2$ = 3,123.00) and standard error (55.88) although both values remain slightly higher than the SLM. The results suggest that the SLM offers a better

explanation of the spatial variation in operational industries and leaves less variation unexplained than the OLS or SEM models.

The goodness-of-fit statistics and the spatial dependence diagnostics in Table 6 jointly suggest that the Spatial Lag Model (SLM) is the preferred specification for this study. The better performance of the SLM indicates that spatial dependence works mainly through interactions among neighboring districts rather than spatially correlated error terms. That is, the industrial activity of neighboring districts has a significant impact on the number of operational manufacturing industries in a district, indicating the presence of spatial spillover effects and industrial agglomeration. Thus, the SLM provides the most appropriate framework to analyse the determinants of industrial operations and spatial structure of manufacturing activity in Maharashtra.

7. Discussion:

The spatial analysis indicates that the industrial activity in Maharashtra is locationally concentrated and not randomly distributed (Chakraborty, 2024; Tewari, 2025). The positive values of Moran's I for operational factories, registered factories, physical working capital, fixed-capital value addition, output, profit and male employment imply that districts with relatively similar levels of industrial activity tend to be located near to each other. The pattern is confirmed by the LISA results that show High–High concentrations in and around Mumbai Metropolitan Region towards Pune and Low–Low concentrations in parts of Marathwada and Vidarbha. This concentration is consistent with the theoretical mechanisms of agglomeration economies, particularly sharing, matching and learning, through which firms benefit from local markets, specialized labour, supplier networks and the exchange of knowledge (Krugman, 1991; Duranton & Puga, 2004). Similar evidence from India also indicates that neighboring effects are important in accounting for the geographical concentration of manufacturing activities, reaffirming the importance of explicitly including spatial dependence in industrial analysis (Tewari, 2025; Azhar & Adil, 2019).

The PCA results complement the evidence on the internal structure of industrial activity. The four components retained explained 95.96% of the total variance and PC1 explained 83.21%. The high loadings of operating factories, employment, wages, working capital, value added, profit, depreciation and net income indicate that the first dimension of variation across the districts of Maharashtra is associated with industrial size, employment and economic performance. The principal features of PC2 are output, fixed capital, plant and machinery, fuel

consumption and capital formation and it represents a different dimension of production and capital investment. This pattern is consistent with the research that emphasizes capital accumulation, productivity and productive capabilities as important dimensions of regional industrial performance (Chang & Andreoni, 2020; Oqubay, 2025; Liu et al., 2025). At the same time, because of the relatively small number of districts relative to the 47 indicators, the PCA should be seen as an exploratory data-reduction procedure rather than a confirmatory factor model (Morton & Altschul, 2019; Kline, 2013; Garson, 2022).

The results of the spatial regression strengthen the evidence that industrial activity cannot be properly understood independently of neighbouring districts. The OLS model explains a large part of the variation in operational industries, but the LM diagnostics support a spatial-lag specification: LM-Lag is significant (8.2898, $p = 0.00399$) and so is Robust LM-Lag (8.1415, $p = 0.00433$), while both LM-Error and Robust LM-Error are insignificant. This result is in line with the basic assumption of spatial econometrics that economic processes are crossing administrative borders and that neighbouring regions can influence each other via production linkages, labour mobility, investment flows and knowledge spillovers (Anselin, 1988; LeSage & Pace, 2009; Elhorst, 2014). Research on industrial agglomeration also shows that neighbouring effects can be a contributing factor to the geographical concentration of manufacturing activity, providing a sound basis for considering spatial dependence in industrial models (Viladecans-Marsal, 2004; Yu et al., 2023).

Additional model comparison statistics also point toward the preference of the Spatial Lag Model. The SLM has a much higher log-likelihood, lower AIC and BIC, and much lower residual variance and standard error than the OLS. It also slightly outperforms the SEM in terms of log-likelihood, residual variance and standard error, although the SEM has marginally lower AIC and BIC in the reported estimates. Therefore, the SLM should not be selected based solely on goodness-of-fit statistics. The model is supported by the combined evidence of the LM diagnostics, the estimated spatial dependence, and the comparative model performance (Anselin et al., 1996; Florax & Graaff, 2004). This is important as spatial interaction is theoretically plausible in Maharashtra's industrial system, where neighbouring districts can share labour markets, transport networks, suppliers, markets and production chains.

The negative spatial autoregressive coefficient, however, should be interpreted cautiously. This means that after consideration of the industrial characteristics included, there is a negative conditional spatial correlation between the number of operating industries in adjacent districts.

It should not be interpreted directly as evidence of competition, since the present model does not explicitly measure competitive displacement, land-price effects, investment diversion or firm relocation. Another more defensible interpretation is that industrial activity may be selectively concentrated in established industrial locations rather than being diffusely distributed across neighbouring districts. This interpretation is consistent with cumulative-causation and path-dependency perspectives, according to which established industrial centers can reinforce their advantages through accumulated infrastructure, skills, markets and production networks (Myrdal, 1957; Martin & Sunley, 2006; Liu et al., 2026). This is also consistent with the more general argument that increasing returns and spatial concentration can lead to persistent differences between core and peripheral regions (Krugman, 1991; Fujita et al., 1999).

Thus, the existence of High–High and Low–Low clusters shows the existence of an uneven pattern of industrial development in the state of Maharashtra rather than a uniform pattern. The concentration of operational industries in the Mumbai–Thane–Raigad–Pune region is indicative of the concentration of industrial capabilities and agglomeration advantages while the continued existence of Low–Low clusters suggests a relatively weak integration into the major industrial networks of the state. Importantly, the analysis here does not directly test for the effects of infrastructure, institutional quality, transport accessibility, or industrial policy. Thus, these factors should be regarded as plausible contextual explanations instead of empirically established determinants in this study. This difference is especially relevant as the location of industry is usually determined by a number of interacting factors such as market access, transport costs, availability of labour, technological capabilities and institutional conditions (Duranton & Puga, 2004; Elhorst, 2014; Iammarino & McCann, 2013).

To summarize, the three analytical approaches provide complementary evidence. PCA identifies the first dimensions of industrial characteristics, Moran's I and LISA identify their spatial concentration and the SLM shows the relevance of neighbouring districts to explain variations in operational industries. The results collectively indicate that the industrial geography of Maharashtra is characterized by combination of industrial scale, capital and employment intensity and spatial concentration. The results imply that rather than a straightforward process of industrial diffusion, there is a path-dependent geography in which established industrial centers enjoy substantial locational advantages. This has implications for regional industrial policy: interventions in lagging districts need to address not only individual industrial inputs, but also their connectivity with wider regional production systems. However,

because the analysis is based on cross-sectional district-level data, causal and endogeneity issues cannot be fully addressed; future research should include infrastructure, institutional and policy variables and test alternative spatial-weight matrices and spatial count-data specifications.

8. Conclusion:

The present research examined the spatial distribution and important features of operational manufacturing industries in the districts of Maharashtra using PCA, Moran's I, LISA, OLS and spatial regression. The results indicate a strong spatial differentiation of industrial activity. The operational industries are concentrated in the established industrial regions and relatively weaker concentrations across parts of Marathwada and Vidarbha. The Moran's I results and LISA analysis indicate that the industrial activity is not randomly distributed, but shows strong spatial clustering especially around the Mumbai Metropolitan Region, Pune and adjoining industrial districts. These findings reinforce the importance of agglomeration economies, accumulated infrastructure, market accessibility, and regional production networks in shaping Maharashtra's industrial geography (Dhulipala & Patil, 2021; Chakraborty, 2024; Rahman & Dua, (2024).

The results have direct policy implications regarding regional industry policy. The persistence of High-High and Low-Low clusters implies that the implementation of a single statewide industry policy would not suffice. There is a need for the targeting of policies geographically, with greater focus on the reinforcement of secondary industry centers, and better connectivity between established industrial clusters and areas less developed industrially. Transportation infrastructure development, establishment of industry centres, skilled labour availability, financial access and regional industrial linkages can increase the potential of peripheral districts in participating in the broader industrial network. However, the high level of industrialization implies that diffusion from established industrial growth centres to neighbouring districts cannot be taken for granted.

Limitations of the study Firstly, the analysis is based on cross-sectional ASI data and therefore cannot establish temporal changes and causal relationships. Second, the number of district observations is relatively small in comparison with the number of original industrial indicators, which puts constraints on the robustness of multivariate analysis. Thus, the PCA was used mainly as a data-reduction technique rather than as evidence of causal determinants. Third, the ASI measures registered or formal manufacturing activity and does not adequately capture

informal industrial activities. Fourth, key contextual determinants such as infrastructure quality, institutional capacity, governance, land policy, and industrial incentives could not be directly included in the current model. Finally, the spatial analysis relies on a particular spatial-weight specification that limits the generalizability of the results to alternative definitions of neighbourhood relationships.

The OLS model explains a large part of the variation in industrial operation, but spatial diagnostic tests show that the assumption of independent distribution of observations is not fulfilled. Industrial activity is spatial in nature and neighbouring districts may have an influence on each other's industrial performance through shared infrastructure, labour markets, supply chains and investment networks. Hence, OLS estimates may be inefficient in the presence of spatial dependence. As the dependent variable is the number of operational industries, count-data models (Poisson or Negative Binomial) could also be considered. However, the present study uses spatial regression models to explicitly capture spatial dependence of districts which is the primary objective of the analysis.

Future research should address these limitations by employing multi-year or panel ASI data, integrating infrastructure, institutional, policy, and accessibility variables, and testing alternative spatial-weight matrices. It should also examine possible endogeneity between industrial scale, output, profit, capital and employment using appropriate instrumental-variable, panel or other causal methods. A comparative study with other major industrial states such as Gujarat and Tamil Nadu would help us to see whether the spatial patterns revealed in Maharashtra are specific to the state or are indicative of broader processes of industrial regionalization in India.

The study finds that the industrial geography of Maharashtra is marked by the presence of industrial agglomeration, regional inequality, inter-related industrial characteristics and spatial dependence. Hence the main contribution of the study is less the identification of individual industrial variables than the demonstration that the functioning of manufacturing industries has to be understood in its wider spatial context. The results give an empirical foundation to move from aggregate state-level industrial planning to district-specific, spatially coordinated and evidence-based industrial policies.

Ethical Approval:

The study utilized secondary data, which is publicly available and free to access through the Directorate of Economics and Statistics, Planning Department, Government of Maharashtra. All

protocols and ethical guidelines of institutional committees were adhered to, and therefore, no ethical approval was required for this data.

Availability of Data and Materials:

The data analysed in this study are derived from the Annual Survey of Industries (ASI) 2023–24 and were obtained from the Commissionerate of Economics and Statistics, Planning Department, Government of Maharashtra. The official ASI information and reports are available through the department's website https://mospi.gov.in/themes/product/65-annual-survey-of-industries

CRediT authorship contribution statement:

Sudeb Pal: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; and Writing - review & editing. **Sumana Bandyopadhyay:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; and Writing - review & editing.

Statements and Declarations:

Competing Interests: The authors have no financial, personal, or professional interests that could be construed as having influenced the research presented in this manuscript.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements:

The author would like to thank the Directorate of Economics and Statistics, Planning Department, Government of Maharashtra, for providing the data from the Annual Survey of Industries.

References:

1. Abdali, Y., & Rajaei, S. A. (2020). Determining the components of physical resilience in the residential texture of Bojnourd City using Moran’s I spatial autocorrelation.

2. Annual survey of industries 2023-24. *Mumbai: Directorate of Economics and Statistics, Planning Department, Government of Maharashtra. (2026).*
3. Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic Publishers.
4. Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis*, *27*(2), 93-115.
5. Anselin, L. (2005). Exploring spatial data with GeoDaTM: a workbook. *Center for spatially integrated social science*, *1963*, 157.
6. Anselin, L. (2022). Spatial econometrics. *Handbook of spatial analysis in the social sciences*, 101-122.
7. Anselin, L., & Smirnov, O. (1996). Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of regional science*, *36*(1), 67-89.
8. Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional science and urban economics*, *26*(1), 77-104.
9. Ashton, W. S. (2009). The structure, function, and evolution of a regional industrial ecosystem. *Journal of Industrial ecology*, *13*(2), 228-246.
10. Azhar, A., & Adil, S. (2019). The effects of agglomeration on socio-economic outcomes: A district level panel study of Punjab. *The Pakistan development review*, 159-176.
11. Baltagi, B. H., Song, S. H., & Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of econometrics*, *117*(1), 123-150.
12. Banerjee, S., Gelfand, A. E., & Carlin, B. P. (2025). *Hierarchical modeling and analysis for spatial data*. CRC press.
13. Barakat, M., Ali, A. H., Madkour, T., & Eid, A. (2024). Enhancing operational performance through supply chain integration: the mediating role of resilience capabilities under COVID-19 pandemic. *International Journal of Logistics Systems and Management*, *47*(2), 161-184.
14. Bhalotra, S. (2002). The impact of economic liberalization on employment and wages in India. *ILO, Geneva*, 121-32.
15. Carroll, M. C., Reid, N., & Smith, B. W. (2008). Location quotients versus spatial autocorrelation in identifying potential cluster regions. *The Annals of Regional Science, 42*(2), 449–463.
16. Chakraborty, M. (2024). Industrial clustering and location in India: Sectoral patterns of investments and employments. *Regional Science Policy & Practice*, *16*(6), 100041.

17. Chang, H. J., & Andreoni, A. (2020). Industrial policy in the 21st century. *Development and change*, *51*(2), 324-351.
18. Chen, W., Shen, Y., Wang, Y., & Wu, Q. (2018). How do industrial land price variations affect industrial diffusion? Evidence from a spatial analysis of China. *Land Use Policy*, *71*, 384-394.
19. Chen, Y. (2024). Reconstruction and normalization of LISA for spatial analysis. *PLoS One*, *19*(5), e0303456.
20. Constantin, C. (2014). Principal component analysis-a powerful tool in computing marketing information. *Bulletin of the Transilvania University of Brasov. Series V: Economic Sciences*, 25-30.
21. Craven, B. D., & Islam, S. M. (2011). Ordinary least-squares regression. *The SAGE dictionary of quantitative management research*, *1*, 224-228.
22. De Siano, R., Sciabolazza, V. L., & Sapio, A. (2020). *Regional resilience to climate and environmental shocks: A spatial econometric perspective*. Springer Nature.
23. Dhulipala, S., & Patil, G. R. (2021). Identification of freight generating industry complexes: A descriptive spatial analysis. *Growth and Change*, *52*(4), 2680-2712.
24. Dismuke, C., & Lindrooth, R. (2006). Ordinary least squares. *Methods and designs for outcomes research*, *93*(1), 93-104.
25. Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. *Regional science and urban economics*, *22*(3), 433-452.
26. Economic Survey of Maharashtra 2023–24. Directorate of Economics and Statistics, Planning Department, Government of Maharashtra. Government of Maharashtra. (2024).
27. Elhorst, J. P. (2014). *Spatial econometrics: From cross-sectional data to spatial panels*. Springer.
28. Florax, R. J., & De Graaff, T. (2004). The performance of diagnostic tests for spatial dependence in linear regression models: a meta-analysis of simulation studies. In *Advances in spatial econometrics: Methodology, tools and applications* (pp. 29-65). Berlin, Heidelberg: Springer Berlin Heidelberg.
29. Fujita, M., & Thisse, J. F. (2002). *Economics of agglomeration: Cities, industrial location, and regional growth*. Cambridge University Press.
30. Garson, G. D. (2022). *Factor analysis and dimension reduction in R: A social scientist's toolkit*. Taylor & Francis.

31. Getis, A., & Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical analysis*, *36*(2), 90-104.
32. Golgher, A. B., & Voss, P. R. (2016). How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography*, *4*(3), 175-205.
33. Hirschman, A. O. (1958). *The strategy of economic development*. Yale University Press.
34. Hua, W., Yuxin, Z., Mengyu, W., Jiqiang, N., Xueye, C., & Yang, Z. (2021). Spatial Characteristics and Driving Forces of Cultivated Land Changes by Coupling Spatial Autocorrelation Model and Spatial-temporal Big Data. *KSII Transactions on Internet & Information Systems*, *15*(2).
35. Iammarino, S., & McCann, P. (2013). *Multinationals and economic geography: Location, technology and innovation*. Edward Elgar Publishing.
36. Isard, W. (1956). *Location and space-economy: A general theory relating to industrial location, market areas, land use, trade, and urban structure*. MIT Press.
37. Jesri, N., Saghafipour, A., Koohpaei, A., Farzinnia, B., Jooshin, M. K., Abolkheirian, S., & Sarvi, M. (2021). Mapping and spatial pattern analysis of COVID-19 in Central Iran using the Local Indicators of Spatial Association (LISA). *BMC Public Health*, *21*(1), 2227.
38. Kalota, D., & Khan, M. F. (2011). Dynamics of Industrialization in Maharashtra: A Spatial Analysis. *Researchgate. net/publication/216680054*.
39. Kitson, M., & Michie, J. (2014). The deindustrial revolution: The rise and fall of UK manufacturing. *Cambridge Journal of Regions, Economy and Society, 7*(3), 315–331.
40. Kline, R. B. (2013). Exploratory and confirmatory factor analysis. *Applied quantitative analysis in the social sciences*, *1*, 171-207.
41. Kollmeyer, C., & Pichler, F. (2013). Is deindustrialization causing high unemployment in affluent countries? Evidence from 18 OECD countries, 1970–2008. *Social Forces, 91*(3), 785–811.
42. Kotwal, A., Ramaswami, B., & Wadhwa, W. (2011). Economic liberalization and Indian economic growth: What's the evidence?. *Journal of Economic Literature*, *49*(4), 1152-1199.
43. Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy, 99*(3), 483–499.
44. LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press.

45. Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School, 22*(2), 139–191. https://doi.org/10.1111/j.1467-9957.1954.tb00021.x
46. Liu, G., Bi, R., Wang, S., Li, F., & Guo, G. (2013). The use of spatial autocorrelation analysis to identify PAHs pollution hotspots at an industrially contaminated site. *Environmental monitoring and assessment*, *185*(11), 9549-9558.
47. Liu, Y., & Cao, J. (2026). Decoding the Formation Mechanisms of Sustainable Industrial Heritage Corridors: The Institution–Network–Cluster Model from Jiangsu, China. *Sustainability*, *18*(8), 3757.
48. Liu, Y., Shen, L., & Ullah, F. (2025). Linking manufacturing smart transformation to regional economic development in China: The Crucial mediation of regional innovation capacity. *Systems*, *13*(5), 389.
49. Luan, Y., Wang, Z., & Ren, J. (2025). The spatiotemporal heterogeneities of deindustrialization in mainland china. *Systems*, *13*(12), 1071.
50. Majumder, S., & Cajee, L. (2026). Unveiling socio-economic inequalities and spatial disparities across districts in Meghalaya: insights from PCA, Moran's I and LISA. *Letters in Spatial and Resource Sciences*, *19*(1), 3.
51. Marshall, A. (1890). *Principles of economics*. Macmillan.
52. Martin, R., & Sunley, P. (2006). Path dependence and regional economic evolution. *Journal of economic geography*, *6*(4), 395-437.
53. Mendez, P., Atienza, M., & Modrego, F. (2023). Urbanization and productivity at a global level: new empirical evidence for the services sector. *Regional Science Policy & Practice*, *15*(9), 1981-1997.
54. Ministry of Statistics and Programme Implementation. (2025). *Annual Survey of Industries 2023–24* (Vol. I). National Statistical Office, Government of India. https://www.mospi.gov.in.
55. Moghimi, F., Baradaran, V., & Hosseinian, A. H. (2023). Identifying the influential factors on the effectiveness of industrial parks and using an MCDM method to rank them: case study of Iran. *Journal of Facilities Management*, *21*(5), 816-844.
56. Morton, F. B., & Altschul, D. (2019). Data reduction analyses of animal behaviour: avoiding Kaiser's criterion and adopting more robust automated methods. *Animal behaviour*, *149*, 89-95.
57. Myrdal, G. (1957). *Economic theory and under-developed regions*. Gerald Duckworth & Co.

58. Naudé, W., & Szirmai, A. (2012). The importance of manufacturing in economic development: Past, present and future perspectives.
59. NITI Aayog. (2024). *Annual report 2023–24*. Government of India.
60. O'Leary, B., Reiners Jr, J. J., Xu, X., & Lemke, L. D. (2016). Identification and influence of spatio-temporal outliers in urban air quality measurements. *Science of the Total Environment*, *573*, 55-65.
61. Oqubay, A. (2025). The theory and practice of industrial policy. *Industrial Policy and Economic Development*, 7-58.
62. Perroux, F. (1955). Note sur la notion de pôle de croissance. *Économie Appliquée, 8*, 307–320.
63. Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review, 76*(6), 77–90.
64. Puntanen, S., & Styan, G. P. (2011). Best Linear Unbiased Estimation in Linear Models.
65. Qu, X., & Lee, L. F. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, *184*(2), 209-232.
66. Qu, Y., Zhang, H., & Liu, X. (2023). Spatial heterogeneity of industrial development and regional economic growth: Evidence from China. *Sustainability, 15*, 1–20.
67. Rahman, M., & Dua, S. (2024). Spatial Organisation and Economic Impact of Manufacturing Sectors in Ludhiana: A Cluster Analysis. *Nagarlok-Quarterly Journal of Urban Affairs*, *56*(4), 509.
68. Rodrik, D. (2013). Unconditional convergence in manufacturing. *The Quarterly Journal of Economics, 128*(1), 165–204. https://doi.org/10.1093/qje/qjs047
69. Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). Industry 4.0: The future of productivity and growth in manufacturing industries. *Boston consulting group*, *9*(1), 54-89.
70. Sarwar, S., Mehmood, T., & Iqbal, M. (2026). A novel statistical framework to address FTIR spectral challenges: Hybrid MARS–PCA/KPCA models for pollutants analysis in honey samples. *Chemometrics and Intelligent Laboratory Systems*, 105625.
71. Shang, Y., Meng, J., & Li, X. (2026). Influence of Urban Compactness on the Supply–Demand Matching of Ecosystem Services—A Case Study of Guanzhong Plain Urban Agglomeration of China. *Forests*, *17*(6), 634.
72. Srivastava, P., Singh, S. P., & Bala, I. (2026). Industrial Growth and Infrastructure Development in India: Progress, Challenges and the Path to Sustainable Development

(SDG9). *International Journal of Humanities, Commerce and Education*, *2*(5), 270-279.

73. Tate, W. L., Ellram, L. M., Schoenherr, T., & Petersen, K. J. (2014). Global competitive conditions driving the manufacturing location decision. *Business Horizons*, *57*(3), 381-390.
74. Tewari, S. (2025). Manufacturing landscape of India: a district-level analysis using economic census. *Journal of Social and Economic Development*, 1-25.
75. Van Neuss, L. (2019). The drivers of structural change. Journal of economic surveys, 33(1), 309-349.
76. Viladecans-Marsal, E. (2004). Agglomeration economies and industrial location: city-level evidence. *Journal of Economic geography*, *4*(5), 565-582.
77. Wang, X., Zhu, Y., Ren, X., & Gozgor, G. (2023). The impact of digital inclusive finance on the spatial convergence of the green total factor productivity in the Chinese cities. *Applied Economics*, *55*(42), 4871-4889.
78. Wang, Y., Liu, Y., Cao, Z., & Zhang, D. (2024). Prediction of contraction channel scour depth: Based on interpretability analysis and PCA-enhanced SVR. *Journal of Hydroinformatics*, *26*(12), 3287-3305.
79. Wheeler, D., Shaw, G., & Barr, S. (2013). *Statistical techniques in geographical analysis*. Routledge.
80. Whitaker, J. K. (1975). The evolution of Alfred Marshall's economic thought and writings over the years 1867–90. In *The Early Economic Writings of Alfred Marshall, 1867–1890* (pp. 1-113). London: Palgrave Macmillan UK.
81. Wu, J., Wei, Y. D., Li, Q., & Yuan, F. (2018). Economic transition and changing location of manufacturing industry in China: A study of the Yangtze River Delta. *Sustainability*, *10*(8), 2624.
82. Yarova, I. Y. (2021). Mechanism of state regulation of the environmental taxation effectiveness in the context of national security.
83. Yu, Z., Xiao, Z., & Liu, X. (2023). Characterizing the spatial-functional network of regional industrial agglomerations: A data-driven case study in China's greater bay area. *Applied Geography*, *152*, 102901.
84. Zhang, Y., Chen, H., & Liu, Z. (2023). Industrial location and urban spatial restructuring: Evidence from manufacturing clusters. *Cities, 137*, 104311.