International Journal of Industrial Engineering and Operational Research (IJIEOR)

Contents lists available at IJIEOR
International Journal of Industrial Engineering and Operational Research
journal homepage: http://ijieor.ir
Volume 5, No. 5, 2023

Spatial Analysis of Housing Prices Using Geographically Weighted Regression: A Case Study of Hedonic Housing Price Data in Isfahan

Erfan Seif [a]

[a] *Department of Industrial and Systems Engineering, Isfahan University of Technology*.

ARTICLE INFO

Received: 2023/10/15

Revised: 2023/11/15

Accept: 2023/12/25

Keywords:

Geographically Weighted Regression, Hedonic Data, Housing Prices, Isfahan, Spatial Analysis.

ABSTRACT

Housing is a fundamental human need that plays a critical role in the economic stability of families. Housing prices have always experienced significant fluctuations, with numerous factors impacting them. Investigating said factors can lead to a greater understanding of the housing market and more accurate planning. The present study investigated the impact of five factors: floor area, percentage of individuals with professional careers in the studied area, number of bedrooms, building age, and whether the property is a villa or an apartment, using data from 100 residential properties registered in the Isfahan city real estate and documents system. The study employed methods: ordinary least squares regression and geographically weighted regression. In addition to highlighting the importance of floor area and percentage of professionals (which stems from the spatial location of the studied area), the findings demonstrate the superiority of geographically weighted regression over ordinary least squares regression in spatial analyses.

1. Introduction

Real estate is considered an essential asset for Iranian investors and families [10,13]. Due to various reasons, real estate has emerged as a popular investment choice. Housing prices have experienced significant growth during the past decade [15, 16]. Housing prices are of particular

[a] Corresponding author email address: erfan.seyf.96@gmail.com (Erfan Seif).
DOI: https://doi.org/10.22034/ijieor.v5i5.150
Available online 12/25/2023

importance from the economic and social perspectives, with affordable, suitable housing being a critical element in determining quality of life [1, 7, 8].

The housing prices in each area of the city vary based on different factors. Over time, rapid expansion and social and economic transformation led to the emergence of new forces that altered inner-city structures and housing prices. Given the extensive fluctuations in housing market supply and demand, effective planning is necessary to improve the quality of residential environments and address related needs. Success in implementing housing planning policies demands understanding the preferences of consumers and their inclinations towards specific housing characteristics. Thus, determining and estimating housing prices is particularly important to planners and decision-makers [5, 12, 14].

Table 1 examines the literature surrounding housing price analysis.

Table 1. A Review of Literature Surrounding Housing Price Analysis

Author(s)	Year	Subject of the Study
[2]	1998	Introduced Geographically Weighted Regression (GWR) for the first time. In this study, he analyzed and evaluated spatial data using inference methods and compared the results of GWR with traditional inference techniques.
[4]	2000	Estimated housing prices using a multilevel model. He investigated how a multilevel approach could be used to assess the spatial dynamics of the housing market, utilizing housing price data from Cardiff.
[3]	2000	Investigated the relationship between accessibility, housing price, and the location of employment in Belfast.
[6]	2005	Investigated the relationship between housing prices and environmental factors in the state of Ohio, USA, using spatial statistics.
[17]	2007	Investigated the impact of urban green spaces on housing prices in China.
[16]	2007	Investigated the spatial distribution of housing prices in Istanbul.
[9]	2010	Investigated the impact of urban rail transportation on housing prices in Bangkok using GWR and a hedonic pricing model.
[11]	2011	Investigated residential house sales in Calgary from 2000 to 2004 using GWR.

Housing prices are among the indicators that the planners cannot entirely control. However, it is somewhat possible to control this indicator by conducting a spatial analysis of the housing market to identify factors impacting price determination and taking effective measures to increase the efficiency of housing plans and programs [18-22].

Generally speaking, linear regression and indicators, such as building density, have been the most common methods for housing price modeling. Nevertheless, the traditional ordinary least squares estimation method cannot accurately reflect all levels of housing price dispersion due to the non-uniform spatial distribution of housing prices [23-27]

Various models that take spatial characteristics into account have been invented to address this issue. One such model that enables the accurate evaluation of factors influencing housing prices is the geographically weighted regression model [27-30]. This model provides superior results compared to the ordinary regression model [30-32]. This study employed ordinary least squares regression and geographically weighted regression to analyze housing prices in Isfahan. The study also aimed to investigate the average housing prices in Isfahan. To this end, the following five indicators were considered: floor area, the nature of the property (apartment or villa), number of bedrooms, construction date, and the percentage of individuals with professional careers, such as doctors and engineers, in the area, based on census data. A total of 100 data groups, each containing data about the five aforementioned indicators, were obtained from the Isfahan city real estate information system. Data analysis was performed using the EViews software, with a conclusion drawn at the end.

2. Ordinary Least Squares (OLS) Regression Model

In contrast to non-linear least squares and methods of increasing the domain of convergence (see [29,30]), the ordinary least squares is the simplest and most commonly used regression model. This method is typically denoted as OLS. The underlying concept of the ordinary least squares method states that model coefficients should take values that lead to the closest matches between the regression model and the observations. In other words, the regression model should demonstrate the least deviation from said observations. In spatial modeling using OLS, it is assumed that the parameters or coefficients of the statistical model remain constant across locations. Thus, the dependent variable estimated by this model corresponds to the entirety of the study area and also estimates a consistent value across different points. This characteristic is recognized as a weakness of this method in spatial modeling. The one-variable simple linear regression model with one variable is as follows:

$$y = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

In this context, y is the dependent variable, x is the independent variable, ε_i is the error or model deviation in estimation, and β_0 and β_1 are the model parameters or coefficients. The OLS statistical model and model coefficient estimation are expressed using the following relations:

$$y = X\beta + \varepsilon \tag{2}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{3}$$

In this context, T is the matrix transpose, $(X^T X)^{-1}$ is the inverse of the variance-covariance matrix, and X is the matrix of independent variables. The OLS regression model coefficients remain constant across the entire space.

3. Geographically Weighted Regression (GWR) Model

The geographically weighted regression model is an expanded form of the general regression framework, with its essence being as follows:

$$y_i = \beta_0(u_i, v_i) + \Sigma\beta_k(u_i, v_i)X_{ik} + \varepsilon_i \qquad i = 1,2,\dots,n \tag{4}$$

In this context, (u_i, v_i) represents the coordinates of the point i in space. $\beta_k(u_i, v_i)$ is a continuous function of $\beta_k(u, v)$ at each i point, X_{i1} to X_{ip} are explanatory variables at point i, and ε_i is the error term. For any given dataset, the $\beta_k(u_i, v_i)$ regional parameters are estimated using weighted least squares steps.

The W_{ij} weights for j=1, 2,…,n at each (u_i, v_i) location are obtained as a continuous function of distances between point i and other data points. Now, consider the following matrix:

$$\begin{bmatrix} \beta_0(u_1, v_1) & \cdots & \beta_p(u_1, v_1) \\ \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \cdots & \beta_p(u_n, v_n) \end{bmatrix} \tag{5}$$

This is the matrix of regional parameters, with each row being derived using the following relation:

$$\hat{\beta}_i = (X^T W(i) X)^{-1} X^T W(i) Y \tag{6}$$

In this context, i=1,2,…,n indicates the matrix rows, X is the matrix of independent variables, Y is the dependent variable, and W(i) is an n×n spatial weight matrix.

As previously stated, W(i) is the spatial weight matrix and is defined as follows:

$$W(i) = diag\lfloor W_{i1}, W_{i2}, \dots, W_{in} \rfloor \tag{7}$$

In this context, W(i) is the weight matrix based on location i, such that observations closer to i have higher weights than those farther from i.

This relation uses a least squares estimator. However, the weight matrix is not constant. Thus, W(i) must be calculated for each point i, and W_{ij} approximates the value of each data point at location i. Data points closer to i have greater weight in estimating β(i) parameters compared to more distant points [25-33].

4. Weighting Method

As previously stated, W_{ij} represents the weight of the j data point at the i-th regression point, and dij represents the distance between the i-th regression point and the j-th data point.

In numerous cases, the Gaussian form of the weight framework is more common. A Gaussian function is as follows:

$$\int_{-\infty}^{\infty} a e^{\frac{-(x-b)^2}{2c^2}}\, dx = ac\sqrt{\pi} \qquad \int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi} \tag{8}$$

In this context, a is the peak height of the curve, b is the peak center position, and c is the bandwidth.

Two types of weight frameworks exist: fixed and adaptive. The present study employs the fixed weight framework, which is calculated as follows for each location i in the regional regression model:

$$W_{ij} = exp\left[-\frac{1}{2}({}^{d_{ij}}/_{h})^2\right] \tag{9}$$

In this context, d_{ij} is the distance between locations i and j, and h is the bandwidth. The increase in h decreases the kernel function slope and includes more data points in the regional measurement. Thus, it is necessary to determine the optimal h value in GWR.

5. Efficiency Criteria for OLS and GWR Regression Models

This study employs the R^2 determination coefficient and Akaike Information Criterion (AIC) to compare the validity of regression models. The R^2 coefficient indicates the percentage of variance in the dependent variable that is explained by the independent variables. This means that it indicates the percentage of total Y variance explained by X independent variables. This numerical value ranges from zero to one. The closer this value approaches one, the more representative it becomes of the dependent variable's variance estimation by the independent variables.

If the standard deviations of X and Y are represented as S_X and S_Y, respectively. Their covariance is denoted as $Cov_{y,x}$, and the determination coefficient can be calculated using the following equation:

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{xy}} \tag{10}$$

The AIC method measures relative efficiency and demonstrates the data loss caused by a statistical model. Essentially, this criterion creates a balance between model accuracy and complexity. The studies by Foudy et al. and Wang et al. shows that a low value of this criterion indicates that the estimated value of the model is closer to the observed value. The corrected Akaike criterion can be calculated using the following equation:

$$AIC_c = \text{AIC} + \frac{2k(k+1)}{n-k-1} \tag{11}$$

6. Datasets

The data employed in this study consists of information from 100 residential properties in Isfahan for the year 2019, extracted from the city's real estate registration system. The data was collected from various city areas with decent dispersion. The information on the following five indicators was gathered for all properties in accordance with data from the Iran Statistical Center: floor area, number of bedrooms, construction date, property type (villa or apartment), and the percentage of professionals, including doctors, engineers, and lawyers.

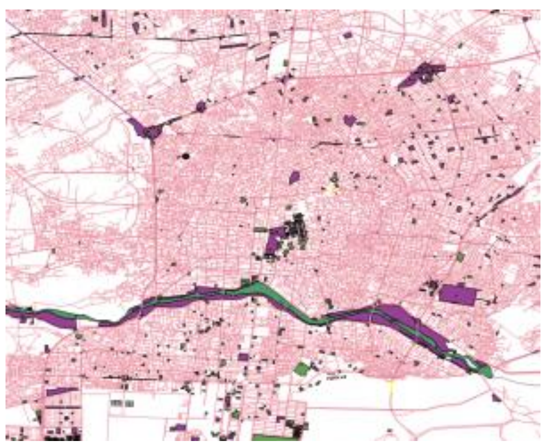

Figure 1. GIS Map of Isfahan, Source: GISPlus Website

7. Applying Geographically Weighted Regression to Investigate Housing Prices

This section of the study investigates ordinary least squares regression and geographically weighted regression models for housing price data in Isfahan. The EViews software was employed for data analysis. The hedonic (binary) variables used in the model are defined as follows:

FLOORSZ: Property floor area in square meters.

BEDS2: 1 for 2 or more bedrooms, 0 for other cases.

BLD80: 1 for construction dates within 2001-2010, 0 for other cases.

BLD90: 1 for construction dates within 2011-2015, 0 for other cases.

BLD95: 1 for construction dates within 2016-2019, 0 for other cases.

TYPEFLAT: 0 for apartment properties, 1 for villa properties.

PROF: Percentage of professionals in the specified area.

The dependent variable is property price, denoted as PURCHASE.

The value and distribution of data are illustrated below.

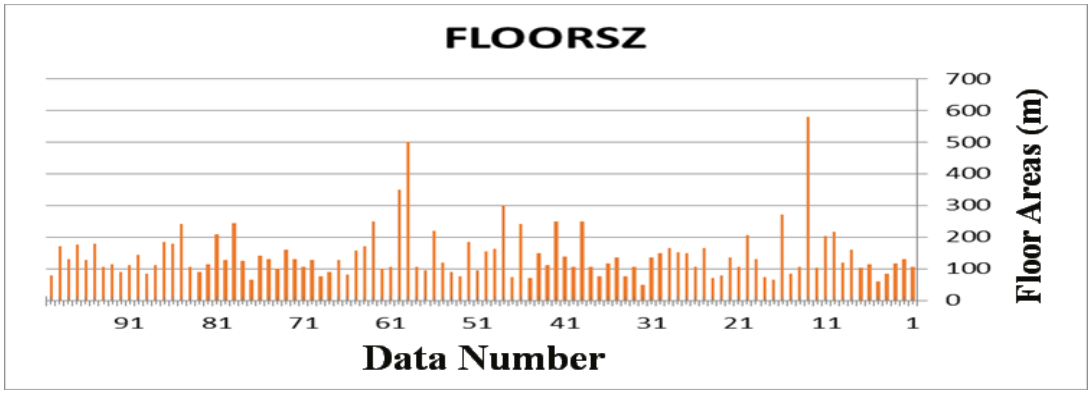

Figure 1. Floor Areas in Used Data, Source: Isfahan Real Estate System

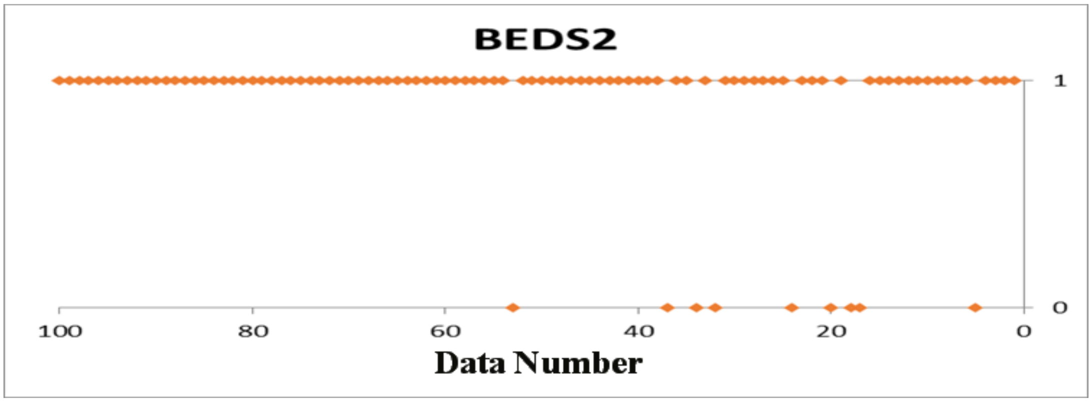

Figure 2. Number of Rooms in Investigated Properties, Source: Isfahan Real Estate System

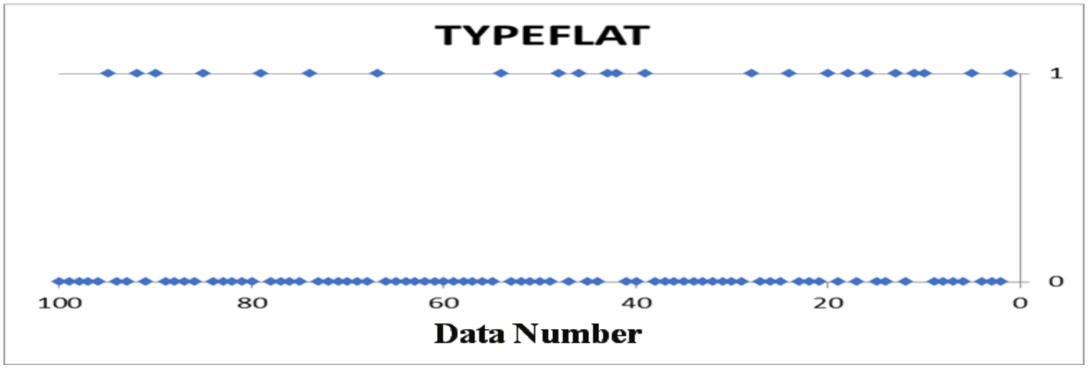

Figure 3. Type of Investigated Properties (Apartment or villa), Source: Isfahan Real Estate System

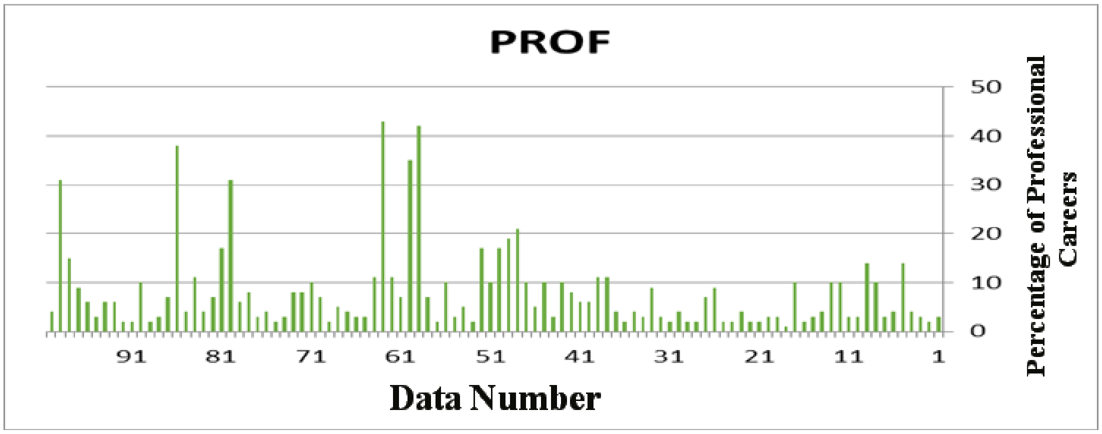

Figure 4. Percentage of Professional Careers in Investigated Areas, Source: Iran Statistical Center

8. Results

The data obtained from the software indicates that the average housing price in Isfahan during the last quarter of 2019 is 110,289,000 Rials. The available data also demonstrates that the PROF and FLOORSZ variables are statistically significant and must not be overlooked.

Table 2. Data Obtained from Software, Source: Author

Prob.	t- Statistic	Std. Error	Coefficient	Variable
0.0393	-2.090436	447.6036	-935.6868	BEDS2
0.7334	0.341698	1223.345	418.0140	BLD80
0.7331	0.342028	1270.802	434.6495	BLD90
0.4839	0.702843	1236.852	869.3126	BLD95
0.0000	9.260010	16.96780	157.1220	PROF
0.0000	9.996358	1.938895	19.38189	FLOORSZ
0.1367	-1.501361	344.2338	-516.8193	TYPEFLAT
0.0992	-1.665801	1303.928	-2172.085	C

A comparison between the data obtained from geographically weighted regression and ordinary least squares regression methods demonstrates the superiority of geographically weighted regression in spatial data analysis.

Table 3. Comparison of OLS and GWR Methods, Source: Author

GWR	OLS	Criterion
0.873211	0.820112	R- squared
0.834126	0.806425	Adjusted R-squared

GWR	OLS	Criterion
16.31293	17.05236	Akaike info criterion
1810	None	Range Length

9. Conclusion

This study conducted a spatial analysis of housing prices in Isfahan using hedonic pricing variables. A dataset of 100 groups, representing five characteristics —floor area, number of bedrooms, construction data, property type, and percentage of professionals in the region —was analyzed using geographically weighted regression with a bandwidth of 1,810 meters and ordinary least squares regression. The findings demonstrate the superiority of geographically weighted regression over ordinary least squares regression in spatial analyses. Additionally, the average housing price in Isfahan during the final quarter of 2019 was calculated at 110,289,000 Rials. It was determined that two variables, floor area and the percentage of professionals in the studied region, are statistically significant.

It is essential to note that quantitative analyses can lead to superior decision-making among individuals during housing selection and reduce investment-related risks. Therefore, private and government organizations related to the housing market can gain better information and improve decision-making through spatial analyses of the real estate market.

Future studies can utilize environmental factors and social conditions as dependent variables in models to measure their impact on housing prices in different areas. The geographically weighted regression model can also be expanded. For instance, considering non-Euclidean distance metrics instead of Euclidean distance would allow researchers to evaluate the impact of distance measurement indices in various Iranian cities with different environmental conditions.

References

[1] Maher, C. (1994). Housing pricing and geographical scale: Australian cities in 1980. Urban Studies.

[2] Fotheringham, A. S., & Charlton, M. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environment and Planning A.

[3] Adair, A., McGreal, A., Cooper, J., & Ryley, T. (2000). House price and accessibility: Testing relationships within the Belfast urban area. Housing Studies.

[4] Orford, S. (2000). Modeling spatial structures in local housing market dynamics: A multilevel perspective. Urban Studies.

[5] Malpezzi, S. (2003). Chapter 5. Hedonic pricing models: A selective and applied review in housing economics and public policy. In Housing Economics and Public Policy.

[6] Brasington, D. (2005). Demand for environmental quality: A spatial hedonic analysis. Regional Science and Urban Economics.

[7] Akbari, N., & Tavasoli, N. (2008). Analysis of the impact of municipal taxes on housing prices; Case study: Isfahan. Quarterly Journal of Economic Studies, 1.

[8] Foroughi, F. (2010). Spatial analysis of factors affecting housing prices in Iran: A geographically weighted regression approach. Master's Thesis, Faculty of Administrative Sciences and Economics, University of Isfahan.

[9] Vichiensan, V., & Miyamoto, K. (2010). Influence of urban rail transit on house value: Spatial hedonic analysis in Bangkok.

[10] Geng, J., Cao, K., Yu, L., & Tang, Y. (2011). Geographically weighted regression model-based spatial analysis of house price in Shenzhen.

[11] Huang, B., & Wu, B. (2011). Geographically and temporally weighted regression for modeling spatio-temporal variation in house price.

[12] Samadi, S., & Moeini, S. (2012). Analysis of metropolitan housing prices and urban growth boundaries in Iran using panel data model in selected cities: Tehran, Isfahan, and Shiraz. Urban and Regional Studies and Research, 14.

[13] Gholipour, H., & Bazrafshan, E. (2012). Determinants of returns on housing.

[14] Erfanian, M., Hosseinkhah, M., & Alijanpour, A. (2013). An introduction to OLS and GWR multivariate regression methods in spatial modeling of land use effects on water quality. Watershed Management Promotion and Development, 1.

[15] Rahnama, M., Asadi, A., & Razavi, M. (2013). Spatial analysis of housing prices in Mashhad using geographically weighted regression. Urban Ecology Research Journal, 7.

[16] Ozus, E. (2007). Spatial analysis of residential prices in Istanbul. [Publication info incomplete].

[17] Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space.

[18] Rafeian, M., & Zahed, N. (2017). Spatial analysis of neighborhood deterioration in Qom using geographically weighted regression. Urban Planning Geography Studies, 2. (Published 2018)

[19] Saremi, H., Heydari, M., & Aghaei, F. (2018). Spatial analysis of housing prices using geographically weighted regression technique; Case study: District 2 of Tehran Municipality. Urban Economics, 2.

[20] Nikpour, A., Rezazadeh, M., & Elhagholi Tabar Nashli, F. (2019). Analysis of factors affecting land price using geographically weighted regression model; Case study: Babolsar.

[21] Lu, B., Charlton, M., & Fotheringham, S. (2011). Geographically weighted regression using a non-Euclidean distance metric with a study on London house price data. International Journal of Geographical Information Science.

[22] Saeedikiya, M., Aeeni, Z., Temiz, S., & Kawamorita, H. (2021). Innovation amplifies growth ambitions of early-stage female entrepreneurs: More in Nordics, less in MENA region. In The emerald handbook of women and entrepreneurship in developing economies (pp. 371-390). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80071-326-020211019

[23] Ashourizadeh, S., Saeedikiya, M., Aeeni, Z., and Poorhosseinzadeh, M. (2022). "Human capital and entrepreneurial career choices of immigrants originating from emerging economies: The liability of foreignness

perspective," in International entrepreneurship in emerging markets, eds V. Jafari-Sadeghi and L. P. Dana (New York, NY: Routledge), 109–128. doi: 10.4324/9781003218357-8

[24] Asgari, A., & Ghaderi, J. Hedonic pricing model for determining housing prices in Iranian urban areas. Economic Research.

[25] Mehri, S., Shafie-Khah, M., Siano, P., Moallem, M., Mokhtari, M., & Catalão, J. P. S. (2017). Contribution of tidal power generation system for damping inter-area oscillation. Energy Conversion and Management, 132, 136-146. https://doi.org/10.1016/j.enconman.2016.11.023

[26] Conticini, C., Filar, J.A. and Ganjehlou, A.N., 2009. A note on price volatility in the Australian electricity market. The Proceedings of ANZIAM, 51, pp.C730-C746. DOI: https://doi.org/10.21914/anziamj.v51i0.2655

[27] Boland, J., Filar, J.A., Mohammadian, G. and Nazari, A., 2016. Australian electricity market and price volatility. Annals of Operations Research, 241, pp.357-372. DOI: https://doi.org/10.1007/s10479-011-1033-x

[28] Eshragh, A., Roosta, F., Nazari, A. and Mahoney, M.W., 2022. LSAR: Efficient leverage score sampling algorithm for the analysis of big time series data. Journal of Machine Learning Research, 23(22), pp.1-36. https://www.jmlr.org/papers/volume23/20-247/20-247.pdf

[29] Yadipour, M., Hashemzadeh, F., & Baradarannia, M. (2017, September). Controller design to enlarge the domain of attraction for a class of nonlinear systems. In 2017 International Conference on Research and Education in Mechatronics (REM) (pp. 1-5). IEEE, doi: 10.1109/REM.2017.8075246

[30] Yadipour, M., Hashemzadeh, F., & Baradarannia, M. (2023). A Novel Strategy of Extending the Domain of Attraction of Affine Nonlinear Systems. Journal of Mathematical Sciences, 276(2), 289-299. https://doi.org/10.1007/s10958-023-06741-2

[31] Shahamipour, A., & Farzanmanesh, R. (2015). Analysis of climatic factors in traditional houses with architectural features of Qajar period in Tabriz. Journal of Natural Sciences Research, 5(17), 20-32.

[32] Arshi, M., Hadi-Vencheh, A., Aazami, A., & Hamlehvar, T. The multiple attribute group decision-making problems with interval-valued intuitionistic fuzzy numbers: A linear programming approach. Journal of Optimization in Industrial Engineering, 38(1).

[33] Soltani, K., Farzinvash, L., & Balafar, M. A. (2023). Trust-aware and energy-efficient data gathering in wireless sensor networks using PSO. Soft Computing, 27, 11731–11754. https://doi.org/10.1007/s00500-023-07856-z