Universiti Teknologi MARA
Cawangan Perlis
Jurnal Intelek

Volume 19 Issue 2 (August) 2024

Application of the ARIMA Model in House Price Index in Malaysia

Muhammad Harith Ikhwan Zamri[1], Muhamad Afizi Rifin[1], Norani Amit[1*]
[1]Kolej Pengajian Perkomputeran, Informatik dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, 70300 Seremban, Negeri Sembilan, Malaysia

*Authors' Email Address: [1]harithi720@gmail.com, [2]afizipizi1985@gmail.com, *[3]norani@uitm.edu.my*

Received Date: 24 May 2024
Accepted Date: 30 May 2024
Revised Date: 24 June 2024
Published Date: 31 July 2024

*Corresponding Author

ABSTRACT

The factors that affecting the escalating price of houses in Malaysia are driven by factors such as population growth, income dynamics, interest rates, and GDP. This phenomenon has notably outpaced the growth of household incomes, thus majorly impacting Malaysians. The study's primary goal is to forecast the housing price index in Malaysia from the best model obtained using Box-Jenkins method. aligning with the 2018-2025 National Housing Policy objectives, utilizing advanced machine learning and time series modeling. The objectives guide the research: to find the best model for predicting house price index in Malaysia using Box-Jenkins Method. Utilizing secondary data from the National Property Information Centre (NAPIC) spanning from 1988 to 2023, the study employed the analytical method of ARIMA. The results favored the ARIMA (1,1,1) model as the best model in predicting housing price indexes. This offers an excellent forecasting model for residential properties towards gaining better understanding of their pricing dynamics and offers potential solutions to the issue of housing affordability for Malaysians.

Keywords: *affordability, ARIMA model, Box-Jenkins method, housing price*

INTRODUCTION

A place to live is a basic necessity for humans. A house provides shelter for individuals or families from rain, heat, and wild animals, as well as a place to rest. In short, it is a place where people can live safely. There are many different types of residences available nowadays including terrace houses, semi-detached houses, apartments, flats, bungalows, and condominiums, with varied architectural designs. The price of a house is related to a person's ability to purchase it. House prices significantly impact buyers' decision to purchase a house and are influenced by several factors such as population, income, interest rate, and Gross Domestic Product (GDP) (San Ong, 2013). Due to rapid economic development, most people from the countryside move to the cities to find better job opportunities and attain higher salaries (Adetunji et al., 2022). This phenomenon leads to an increase in demand for houses and consequently affecting housing price.

E-ISSN 2231-7716
DOI: http://10.24191/ji.v19i2.26615

Malaysia is not exempted from the issue of increasing housing price. House buyers are facing the glaring issue of imbalance between the rate of household income and housing prices (Latif et al., 2020). In other words, the price of houses does not match the average salary of Malaysians. As such, houses in Malaysia have become unaffordable. Generally, a house is deemed inexpensive if it costs less than 30% of the gross household income (Olanrewaju & Woon, 2017). However, this is not the case in Malaysia. Therefore, this investigation identifies the need to forecast Malaysia's housing price index. The significant increase in registered unit releases in Malaysia, which exceeds figures recorded in the previous year, suggests a potential recovery in the real estate sector following the pandemic.

Despite a lower total number of unit releases compared to pre-pandemic levels, the upward trend indicates a degree of market recovery and resilience. According to the National Property Information Centre (NAPIC) Press Release for Malaysia Property Market Report H1 2022 (2023), the average sales performance of 36.0% indicates a moderate response to the new launches and the need for a more nuanced understanding of buyer behavior, economic conditions, and other influencing factors. Although there is a clear demand for housing, the challenge lies in the high prices which act as a deterrent for potential buyers. This not only results in a surplus of unsold houses, but also poses a potential economic concern for the future. The mismatch between demand and affordability could lead to adverse economic implications, highlighting the necessity for strategic interventions to address pricing dynamics in the real estate market. Therefore, this study sets out as a guideline that can be used by the Government in projecting housing market trends according to the 2018-2025 National Housing Policy (Jabatan Perumahan Negara, 2018). The development of the National Housing Policy is meant to assist people who are struggling to purchase a home.

LITERATURE REVIEW

There are various factors affecting the determination of house prices including house size, square foot, number of rooms and floors, as well as location. Yahya et al. (2020) stated that the size of the house has the strongest relationship with house price. This finding was reinforced by Zainal et al. (2019) who assert that house prices vary based on the condition or size of the dwelling. Earlier research utilized data from various house types including terrace houses, semi-detached houses, clustered houses, and townhouses, all of which are priced differently according to their size.

Meanwhile, Ramli et al. (2022) found a high correlation between house prices and the factors of inflation rate, unemployment rate, and Gross Domestic Product. Likewise, Jehani et al. (2020) established a long-term relationship between Gross Domestic Product, interest rate, inflation rate, population, and unemployment rate with house prices. Moreover, additional investigations conducted by earlier researchers underscore the significance of the aforementioned factors as crucial elements influencing the escalation or decline of property prices in Malaysia. Latif et al. (2020) investigated the variables that affect the fluctuations of housing prices in Malaysia including Gross Domestic Product, interest rates, unemployment, and inflation.

Auto-Regressive Integrated Moving Average (ARIMA) is extensively employed for forecasting future values, given its status as one of the most prominent time series methods for predicting upcoming trends. This assertion is substantiated by numerous prior studies that have employed the ARIMA method as a reliable approach for predicting future outcomes. Kaur et al. (2023) forecasted the capabilities of ARIMA as highlighted in the areas of environment, health, and atmosphere. ARIMA can provide accurate predictions for future events, especially in the field of time series analysis and forecasting.

ARIMA is a widely adopted method for forecasting future House Price Index as demonstrated by Saudin et al. (2020). Mangaleswaran and Vigneshwari (2020) employed the ARIMA model to make real-time predictions whilst Boitan (2016) used the ARIMA model to analyze the trends of residential property prices in several selected EU nations. All the authors expressed satisfaction with the analysis and stated that ARIMA can be a better method for forecasting. Finally, Alkali (2020) examined real estate price forecasting in Nigeria using ARIMA to obtain an efficient forecasting model for residential properties. It was found that the ARIMA (4,1,1), ARIMA (1,1,1), ARIMA (8,1,1), and ARIMA (1,1,6) models were able to project the price of houses for the next five years. The author stated that ARIMA offers an excellent forecasting model for residential properties towards gaining better understanding of their pricing dynamics.

METHODOLOGY

Box-Jenkins ARIMA Model

The linear structural model called the Box-Jenkins technique utilizes both the stochastic components of a variable and all its historical values to forecast future values for all upcoming periods. This methodology utilizes partition data to separate the data into two parts which are estimation and evaluation. An estimation is used to identify and estimate the parameters of the ARIMA model. The process involves selecting an appropriate order for the autoregressive (AR), differencing (I), and moving average (MA) components. The estimation dataset is crucial for fitting the model and determining the values of these parameters to make accurate predictions. An evaluation dataset assesses the forecasting values by performing various measurement errors. In addition, the partition data is divided into a ratio of ¾ for estimation and ¼ for evaluation. Therefore, based on data obtained, ¾ of the data covers the period from the first quarter of 1988 to the third quarter of 2012, while the remaining ¼ covers the period from the fourth quarter of 2012 to the second quarter of 2023.

The linear structural model called the Box-Jenkins technique utilizes both the stochastic components of a variable and all its historical values to forecast future values for all upcoming periods. In understanding the forecasting capability of the Box-Jenkins technique, there is a need to first understand how time series data are generated for a variable. The process involves selecting an appropriate order for the autoregressive (AR), differencing (I), and moving average (MA) components. The Autoregressive Integrated Moving Average (ARIMA) entails a procedure of data generation that combines the AR and MA processes (p, q), with "I" denoting the series integration order or the differencing operations number ('d') that is necessary to transform the variable's time series into stationary. The ARIMA model (p,d,q) is written as follows:

$$w_t = \mu + \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \ldots + \varphi_p W_{t-p} - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} + \varepsilon_t$$

where;
w_t represent the first difference of the series and is assumed stationary,
Φ_p is the parameter of the autoregressive (AR) part of the model
θ_q is the parameter of the moving average (MA) part of the model
ε_t is the error term at time t.

Model Identification

Model identifications are proposed based on the Autocorrelation Coefficients (ACF) and Partial Autocorrelation Coefficients (PACF) were generated using the correlogram which aids in the determination of the proper ARIMA (p, d, q) values. Table 1 shows the ACF and PACF characteristics in general (Mukopi, 2012).

Table 1: Identification of the Order

Model	ACF	PACF
AR	Spikes decays towards zero	Spikes cut-off to zero at lag p
MA	Spikes cut-off at lag q	Spikes decay to zero

The second strategy that makes use of the statistical test procedure is based on the Augmented Dickey-Fuller (ADF) approach (Darne´ & Diebolt, 2005), which is presently the most widely utilized in practice. The ADF test was employed to check the presence of a unit root. The result of the ADF test is deemed stationary if the p-value is below the significance threshold value.

Model Validation

A variety of indicators, including the Akaike Information Criterion (AIC) and the Bayesian Information Criterion, are used to evaluate and select the best model (Tihi & Popov, 2023). According to Bozdogan (1987), the AIC estimates the prediction error and allows for model selection. When the dataset is relatively tiny, the AIC may choose overfitted models (Tihi & Popov, 2023). Aside from that, compared to the AIC, the BIC fines more estimated model parameters. The resulting model possesses less parameters compared to the AIC when the model selection used minimal BIC. Parsimony indicates that the BIC is better for model selection than AIC. The model is indicated to have a better fit if the BIC value is low (Agyemang et al., 2023). The estimated model's accuracy is subsequently ensured by conducting the Box-Pierce and Ljung-Box tests which distinguish the residuals from white noise. The residuals have random distribution; a higher *p*-value compared to the significance threshold value means that the null hypothesis is non-rejectable.

ANALYSIS AND RESULT

Historical data is required for predicting future events. Therefore, this study utilized secondary data. The sources of secondary data were obtained from the National Property Information Centre (NAPIC) and Jabatan Penilaian dan Perkhidmatan Harta (JPPH). The data ranged from quarter one of year 1988 until quarter two of year 2023 to achieve the study's objectives. This research focused on the housing price index in Malaysia, and the data consisted of quarterly observations.

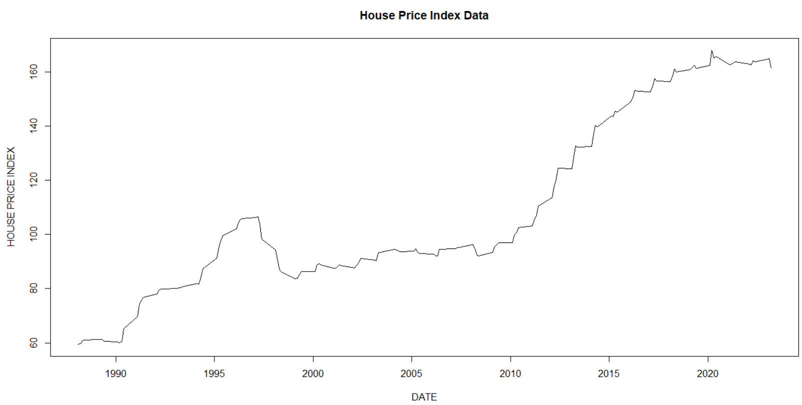

Figure 1: Malaysia House Price Index

Figure 1 shows an upward movement happened from 1988 until 1998, and then the data experienced a downward movement until 2000. The data then began to rise until 2010 and continued to rise dramatically until 2023. Hence, the data for the estimation part was chosen from 1988 (first quarter) to 2012 (third quarter), whilst the data for the evaluation part was picked from 2012 (fourth quarter) to 2023 (second quarter). However, for the purposes of this research, the estimation part was used to identify and estimate the best ARIMA model based on various statistical testing.

In the Box-Jenkins method, the initial part consists of the Unit Root test which was used to establish the stationarity of a time series. The unit root test could be verified using the Augmented Dickey-Fuller (ADF) test. To achieve the stationarity of the dataset, the value of the Augmented Dickey-Fuller (ADF) test must be below the p-value, which is 0.10. It is necessary to perform a second differencing if the data is not stationary following the first one.

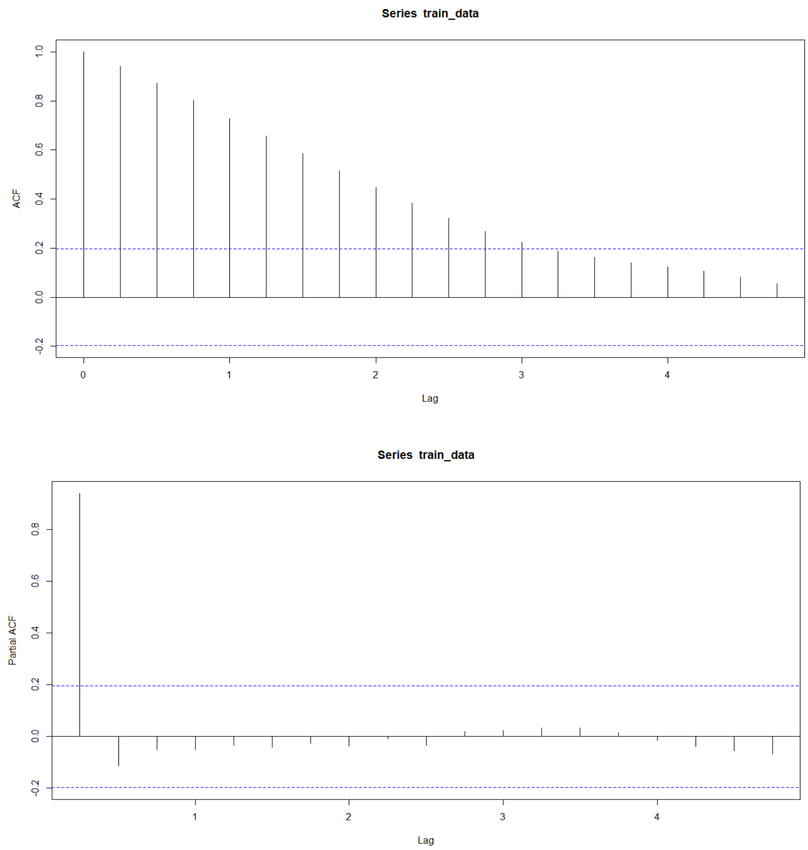

Figure 2: Original ACF and PACF of House Price Index

According to Figure 2, the ACF graph depicts a drop to zero, or in other words, a ‘decay’ behavioral pattern. As a result, the series cannot be said to be stationary. There is just one significant spike in the PACF graph, while the others are insignificant.

Table 2: Augmented-Dickey-Fuller Test for Initial Data

Dickey-Fuller	-2.5759
P-value	0.3382

In this study, the unit root test was carried out and the result indicates non-significance with a probability value of 0.3382, surpassing the threshold of 0.10 as shown in Table 2. As a result, the data is deemed non-stationary. Consequently, the first differencing was implemented, and the results’ subsequent analysis is displayed in Figure 3 and Table 3.

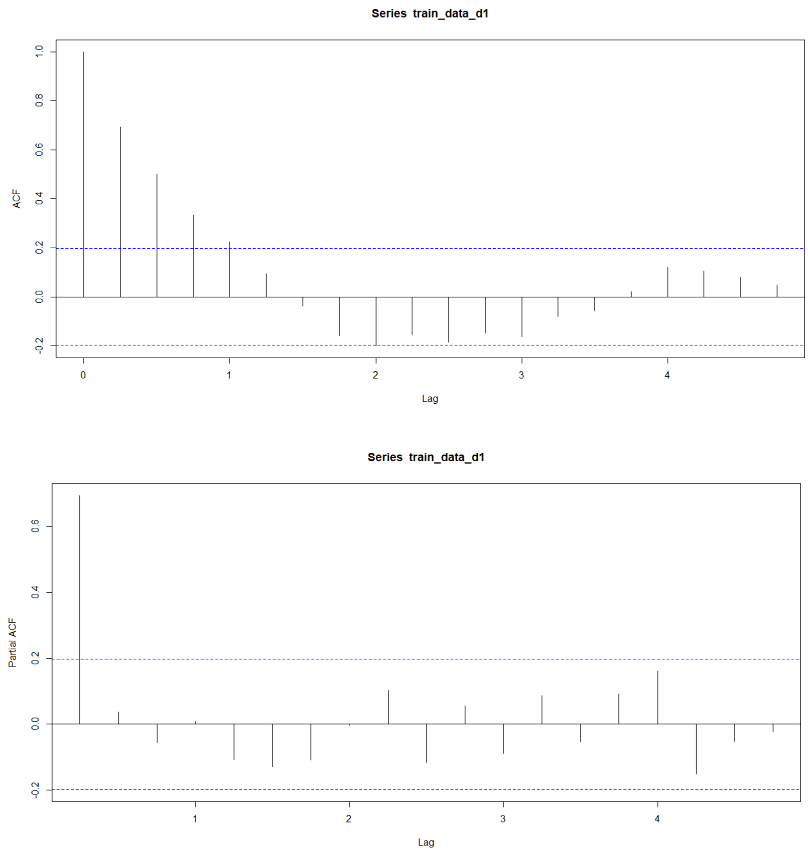

Figure 3: ACF and PACF of House Price Index for First Order of Differencing

Figure 3 indicates that the ACF graph decays at a significantly faster rate when the auto-correlation values shift from positive to negative. For the PACF graph, there is only one significant spike, and the rest are insignificant, same as the original series.

Table 3: Augmented-Dickey-Fuller Test for First Differencing

Dickey-Fuller	-3.1736
P-value	0.09639

The analysis presented in Table 3 shows that the ADF test is deemed stationary as the p-value is 0.09639, i.e., below the significance threshold of 0.10. Next, the number of AR and MA models were determined using the ACF and PACF. Figure 3 illustrates the correlogram which shows that the data is already stationary at the level. The graph autocorrelation and partial correlation are referred to as ACF and PACF, respectively. Because there is only one spike on the PACF graph, it can deduce that AR is 1. Meanwhile, six spikes on the ACF graph indicate that MA is 6. ARIMA (1,1,6) is the only ARIMA model obtained from the correlogram. Based on the values of the ACF and PACF, six models were chosen based on the significant spikes namely ARIMA (1,1,6), ARIMA (1,1,5), ARIMA (1,1,4), ARIMA (1,1,3), ARIMA (1,1,2), and ARIMA (1,1,1). Thus, comparing a few models is needed as it helps in selecting the best model that fits the data well and allows to strike a balance between goodness-of-fit and model complexity (parsimony).

Table 4_presents the summary results of ARIMA model. The process of determining the best model involves comparison between the diagnostics testing results; AIC, BIC and Box L-Jung test. In general, considering that the value of the Box-Ljung test is greater than 0.10, it can be concluded that none of the models exhibit autocorrelation in the residuals. To determine which model is the most suitable, the one with the lowest AIC and BIC value was selected.

Table 4: Summary Results of ARIMA Model

ARIMA Model	Statistical Test		
	AIC	BIC	BOX-LJUNG TEST
ARIMA (1,1,6)	339.0546	359.6523	0.9995
ARIMA (1,1,5)	337.7536	355.7766	0.9994
ARIMA (1,1,4)	337.0017	352.4499	0.9581
ARIMA (1,1,3)	335.0601	347.9337	0.9349
ARIMA (1,1,2)	333.1845	343.4834	0.9329
ARIMA (1,1,1)	**331.5034**	**339.2275**	**0.8544**

Table 4 show the results for each ARIMA models. For the ARIMA (1,1,2) model, the AIC is computed at 333.1845 and BIC at 343.4834, whilst the Box-Ljung test is at 0.9329. Shifting our focus to the ARIMA (1,1,3) model, observe an AIC of 335.0601 and BIC of 347.9337, and Box-Ljung Test of 0.9349. Next, the value AIC, BIC and Box-Ljung test of ARIMA (1,1,4) are 337.0017, 352.4499 and 0.9581, respectively. Meanwhile, ARIMA (1,1,5) calculated the values of AIC is 337.7536, BIC is 355.7766 and Box-Ljung is 0.9994. Notably, the ARIMA (1,1,6) model exhibits the highest values across all parameters, with AIC, BIC, and Box-Ljung test at 339.0546, 359.6523, and 0.9995, respectively. The best model examined showed the lowest values of AIC and BIC. Resultantly, the best model is ARIMA (1,1,1), which meets all the requirements.

CONCLUSION

This study primarily aims to find the best model for predicting the housing price index in Malaysia. Based on the findings, ARIMA (1,1,1) was determined as the best model based on the AIC and BIC value comparison. The AIC and BIC values for ARIMA (1,1,1) has the lowest values among other ARIMA model. Therefore, this model would help in making informed decision as it can be used for forecasting or further analysis. Future studies are advised to add more samples per year to increase the accuracy of the forecasting model. Moreover, future research is recommended to expand the scope when selecting the study variable. More variables should be added instead of only one to give more insights into the housing price problem. Lastly, housing developers need to have a clear insight into homebuyers' expectations and housing preference.

ACKNOWLEDGEMENTS

This report and its analysis would not have materialized without the invaluable support of all the informants. We would like to express our profound appreciation to Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban for consistently encouraging and fostering the development of our writing skills, with an emphasis on producing work of good quality and excellence.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHORS' CONTRIBUTION

All authors provided critical feedback and helped shape the research, analysis and manuscript.

CONFLICT OF INTEREST DECLARATION

We certify that the article is the Authors' and Co-Authors' original work. The article has not been previously published and is not under consideration for publication elsewhere. This research/manuscript has not been submitted for publication nor has it been published in whole or in part elsewhere. We testify to the fact that all Authors have contributed significantly to the work, validity and legitimacy of the data and its interpretation for submission to Jurnal Intelek.

REFERENCES

Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2022). House price prediction using random forest machine learning technique. *Procedia Computer Science*, *199*, 806-813. https://doi.org/10.1016/j.procs.2022.01.100

Agyemang, E. F., Mensah, J. A., Ocran, E., Opoku, E., & Nortey, E. N. (2023). Time series-based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints. *Heliyon*, *9*(12). https://doi.org/10.1016/j.heliyon.2023.e22544

Alkali, M. A. (2020). *Real estate forecasting model for residential market in nigeria* (Unpublished doctoral dissertation). Universiti Teknologi Malaysia.

Boitan, I. A. (2016). Residential property prices' modeling: evidence from selected european countries. *Journal of European Real Estate Research.*

Bozdogan, H. (1987) Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika.* 52(3), 345–370. https://doi.org/10.1007/BF02294361

Darne´, O., & Diebolt, C. (2005). Non-stationarity tests in macroeconomic time series. *New trends in Macroeconomics*, 173–194. https://doi.org/10.1007/3-540-28556-3_9

JabatanPerumahanNegara. (2018). *Dasar perumahan negara (2018-2025).* Jabatan Perumahan Negara Kementerian Pembangunan Kerajaan Tempatan.

Jehani, N. A., Mastani, N. A., Saudin, S., & Ab Malek, I. (2020). A study on the relationship between house price index and its determinants in Malaysia. *Malaysian Journal of Computing (MJoC)*, *5*(2), 515-522. 10.24191/mjoc.v5i2.9466

Kaur, Jatinder, Parmar, Singh, K., Singh, & Sarbjit. (2023). Autoregressive models in environmental forecasting time series: A theoretical and application review. *Environmental Science and Pollution Research*, 1–25. https://doi.org/10.1007/s11356-023-25148-9

Latif, Abdul, N. S., Muhammad, R. K., Nabilah, R., & Khadijah, S. S. (2020). Factors affecting housing prices in malaysia: A literature review. *International Journal of Asian Social Science*, *10*(1), 63–67.

Lwesya, F., Kibambila, V. (2017) A comparative analysis of the application of seasonal ARIMA and exponential smoothing methods in short run forecasting tourist arrivals in Tanzania. *European Journal of Business and Management.* 9(10), 56-69. 10.5958/2320-3226.2020.00001.6

Mangaleswaran, S., & Vigneshwari, S. (2020). Prediction of housing prices using machine learning, time series arima model and artificial neural network. In *Icdsmla 2019: Proceedings of the 1st international conference on data science, machine learning and applications* (pp. 1002–1008). https://doi.org/10.1007/978-981-15-1420-3_110

Mukopi, M. S. (2012). *Prediction of housing prices: An application of the Arima model* (Doctoral dissertation, University of Nairobi, Kenya).

Olanrewaju, A., & Woon, T. C. (2017). An exploration of determinants of affordable housing choice. *International Journal of Housing Markets and Analysis, 10*(5), 703-723.

Press Release Malaysia Property Market Report H1 2022, N. (2023, March.). *Press release malaysia property market report h1 2022.* Valuation and Property Services Department Ministry of Finance Malaysia.

Ramli, Fazilah, Zainal, Rozlin, Kasim, & Rozilah. (2022). Prediction supply for high-cost multi-storey house toward the development of sustainable cities. *International Journal of Sustainable Construction Engineering and Technology*, *13*(2), 100–109.

San Ong, T. (2013). Factors affecting the price of housing in Malaysia. *J. Emerg. Issues Econ. Financ. Bank*, *1*, 414-429.

Saudin, Sharmila, Malek, A., Isnewati, Jehani, Ashakirin, N., & Amaelya, N. (2020). Modelling of Malaysia house price index.

Tihi, N., & Popov, S. (2023). Selection of the best arima models for urban drought prediction. *Fresenius Environmental Bulletin*, 32(6), 2564–2572.

Yahya, N., Zainuddin, N. M. M., Sjarif, N. N. A., & Azmi, N. F. M. (2020). Correlation analysis of factors affecting the prediction of price of terrace houses in Penang, Malaysia: A case study. *Open International Journal of Informatics*, *8*(2), 18-39. https://oiji.utm.my/index.php/oiji/article/view/37

ZZainal, R., Ramli, F., Manap, N., Ali, M., Kasim, N., Noh, H. M., & Musa, S. M. S. (2019). Price prediction model of demand and supply in the housing market. In *MATEC Web of Conferences* (Vol. 266, p. 06015). EDP Sciences.https://doi.org/10.1051/matecconf/201926606015