House Price Indices Prediction By ARIMA, ETS And ARIMA-Xreg

Jiaxuan Gu *

Faculty of Social and Historical Sciences, University College London, London, United Kingdom

* Corresponding Author Email: ZCTPJG1@ucl.ac.uk

Abstract. Residential Property Prices Indices (RPPIs) of United Kingdom is important for policymakers, businesses, investors and individuals because it helps research on the UK's economic direction, develop effective government policies, and guide financial decisions. But there is a lack of papers related to prediction of this factor. Thus, this paper aims to find a model which accurately and comprehensively forecasts RPPI of United Kingdom. This study employs quarterly data of RPPIs for United Kingdom spanning from the first quarter of 1970 to the second quarter of 2025, which has been seasonally adjusted. ARIMA, ARIMA with external regressors (ARIMA with xreg), and ETS are employed in this study for modelling forecasting of RPPI of United Kingdom. By comparing overall MAE and overall RMSE, ETS outperforms both ARIMA and ARIMA with xreg. This study provides important references for investors, policymakers, and individuals to select appropriate forecasting models for predicting housing prices for their specific purpose, like investment.

Keywords: ARIMA; ETS; House Price; UK.

1. Introduction

Accurate forecasting can help people get access to the future trajectory of key factors, so policymakers or other decision makers can develop effective strategies (such as policies) and mitigate risks.

Owing to the importance of prediction, there are many studies focusing on this field. Zamri, Rifin, and Amit adopted the ARIMA model for prediction of quarterly House Price Index data of Malaysia from the first quarter of 1988 to the second quarter of 2023 [1]. Bełej predicted the dynamics of housing prices during Covid-19 by ARIMA combined with TRAMO/SEATS, the forecasting result shows that the COVID-19 did not lead to a significant decline in Poland’s housing market [2]. Based on the data of CPI, exchange rates, and the number of Albanians traveling abroad from the first quarter of 2001 to the fourth quarter of 2016, Gjika, Puka, and Zaçaj adopted two approaches to forecast Albania’s CPI [3]. The results represent that the SARIMA model outperformed both the ETS model and the hybrid regression model in short-term forecasting [3]. Based on the time series of Turkey's housing sales from January 2008 to April 2018, Ayşe, Melek and Günay utilized the ARIMA model, LSTM model, and a hybrid model combining the two for time series forecasting and evaluated the performance of these models by comparing MAE, MSE, RMSE, and MAPE [4]. They found that the hybrid model (particularly the combination of LSTM with 1500 epochs and ARIMA (1,1,1)) performs the best [4]. Zhanao Sun evaluated the forecasting performance of ARIMA and ETS models on the S&P 500 closing prices (2005-2014) [5]. Since ARIMA has lower values across a suite of error measures, like ME and RMSE, the research identified ARIMA (2,1,3) as the optimal model, validating its effectiveness for short-term trend prediction and its potential as a tool for investment guidance [5]. RPPI of United Kingdom is essential since it serves as an important tool for researching on the UK's economic direction, developing effective government policies, and guiding financial decisions for both businesses and individuals. However, there is a lack of papers on forecasting the RPPI of United Kingdom.

Therefore, this study seeks to find a model which can produce accurate and comprehensive forecasts of the UK's Residential Property Price Index (RPPI). This study employs quarterly data of RPPIs for United Kingdom from the first quarter of 1970 to the second quarter of 2025, which has been seasonally adjusted. Three approaches that are ARIMA, ARIMA with external regressors

(ARIMA with xreg), and ETS were evaluated for forecasting the UK's Residential Property Price Index. The ETS model demonstrates superior performance, achieving the lowest overall Mean Absolute Error and overall Root Mean Square Error among all evaluated models. Accurate forecasts of the Residential Property Price Index (RPPI) can support policymakers in England in formulating effective policies and assist businesses, investors and individuals to make profitable decisions, while also enabling both supply and demand sides in the housing market to adjust their strategies and achieve supply-demand balance.

2. Data

Residential Property Prices Indices (RPPIs), alternatively known as House Price Indices (HPIs), are index numbers that measure the prices of residential properties over time. This study employs quarterly data of RPPIs for United Kingdom spanning from the first quarter of 1970 to the second quarter of 2025, which comes from OECD [6]. This set of data has been seasonally adjusted. The RPPIs over the sample period are shown in Fig. 1 below.

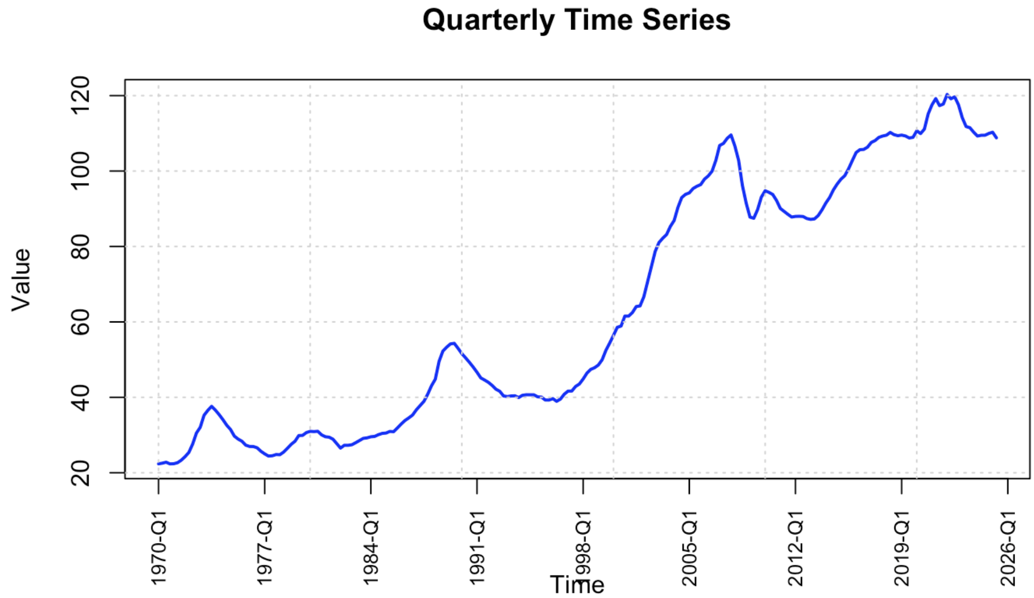

Fig. 1 RPPIs

Fig. 1 shows the trend of the change of RPPIs from the first quarter of 1970 to the second quarter of 2025. This quarterly time series chart shows an overall upward trend in the value over time. There's no obvious seasonal pattern. The period spanning the early to mid-1970s was marked by a noticeable increase that was subsequently followed by a decline. Then, from the mid-1980s to the early 1990s, it experienced a rapid increase, after which it declined and was stable for a period. From the late 1990s onwards, there was a significant and prolonged rapid upward trend, with a peak around 2020. After that, it slightly decreased but remained at a high level. Table 1 presents the basic descriptive statistics of the data.

Table 1. Descriptive Statistics

	mean	SD	max	min	kurtosis	skewness
RPPIs	63.77171	33.03275	120.3031	22.35153	1.448102	0.3008412

Firstly, Table 1 represents that the maximum value of the data is 120.3031, the minimum value is 22.35153, which indicates that this set of RPPI data has a considerable range. This relatively large standard deviation (33.03275) shows that the data distribution is relatively dispersed around the mean, meaning there is a relatively significant degree of variability in the RPPI values. The kurtosis is 1.448102, less than 3, so the distribution of the data is relatively flatter than a normal distribution, with fewer extreme values in the tails. A positive skewness value (0.3008412) represents that the distribution of the data is slightly skewed to the right.

3. Method

3.1. ARIMA

ARIMA models are particularly suitable for forecasting non-seasonal time series because of their sophisticated structure that comprehensively addresses the modeling of non-stationary data [7]. The model includes three fundamental components that are Autoregressive (AR) component, Moving Average (MA) component and Differencing (I) component. Differencing is utilized to ensure that the data is stationary, which is a basic prerequisite for both AR and MA models. Indeed, it will finally return the forecasting values of the stationary time series data to the prediction values of the non-stationary time series data by integrating. The AR captures linear relationships between current values and past observations. Moving Average (MA) helps identify the correlation between the past random error terms of this variable and its current values. It focuses on the fluctuation correlations in stationary sequences that cannot be explained by the observed values themselves, which compensates for the limitation of AR. Therefore, the ARIMA model has advantages in analyzing this set of data and its sequence laws more comprehensively and bringing more accurate prediction. The strength of ARIMA also lies in its clear interpretability of parameters and its ability to fully analyze the inherent dynamics of the data without relying on external variables, which is shown by the comparison of ARIMA and LSTM models in the study of Siami-Namini, Tavakoli, and Namin [8]. The formula of this model is as follows:

$$Y_t = a + \sum_{i=1}^{m} \omega_i Y_{t-i} + \sum_{i=1}^{n} \lambda_t \epsilon_{t-i} + \epsilon_t \tag{1}$$

In this formula, Y_t is Residential Property Prices Indices of United Kingdom. ϵ_t is the error term at time t. a is a constant term. ω_1 to ω_p are the parameters of the AR model, which presents how the value of the past m time affects current value. λ_1 to λ_q are the parameters of the MA model. The relationship between the current value and the error term at n time points in the past is captured by λ_q.

3.2. ARIMA with external regressors (ARIMA with xreg)

For a set of long-term time series data, there are always some unexpected shocks and structural breaks, like natural disasters. This may lead to outliers, which deviate significantly from the overall trend, seasonality, or regular fluctuations, so outliers can distort the accuracy and understanding of some models of the underlying trend and seasonality. Bógalo, Llada, Poncela, and Senra have verified outliers exert a distorting effect on seasonal models [9]. Therefore, there will be a poor fit to history and an unreliable forecast. To alleviate this problem, the ARIMA with xreg model is built. Firstly, time series decomposition is used to identify outliers in the series. Secondly, there are dummy variables created for each outlier to serve as xreg. For example, there is step dummy that stays 1 after the outlier for step-type outliers, or the dummy that has a value of 1 at the time point where the outlier occurs, and 0 otherwise. Finally, these dummy variables are included as xreg in the ARIMA with xreg model. This means the model explicitly quantifies the effect of outliers via estimated coefficients for the outlier variables. This isolates the effect of outliers from the series' intrinsic patterns, allowing the model to more accurately capture underlying trends, seasonality, and autocorrelations. Therefore, the ARIMA with xreg model extends the ARIMA model by adding external regressors (xreg) to resolve problems caused by outliers, improving fitting and forecasting performance. The formula for this model is shown below:

$$Y_t = a + \sum_{i=1}^{m} \omega_i Y_{t-i} + \sum_{i=1}^{n} \lambda_i \epsilon_{t-i} + \sum_{j=1}^{r} \beta_j X_{j,t} + \epsilon_t \tag{2}$$

Xt is a vector composed of dummy variables corresponding to all outliers detected by time series decomposition. Each entry in the Xt vector may be different dummy variables (e.g., step dummy) according to the type of the outliers (e.g., Additive Outliers, Level Shift, Temporary Change). βj represents how the outliers affect the current values. Others are the same as formula 1.

3.3. ETS

The ETS model is a time series forecasting method that decomposes a set of time series data into its core components-error term, trend term, and seasonality term-to analyze the time series and make prediction. The error term shows the random fluctuations that cannot be explained by trend and seasonality. The trend term represents the long-term direction of the set of data. The seasonality term describes repeating patterns occurring at fixed time intervals. The ETS model has a clear structure where each component is well defined, so it is transparent and easily interpretable. Because the error component has two types of additive or multiplicative errors, there can be a total of 30 ETS models [10]. This is the formula of ETS (M, Ad, N), which is the ETS model of this set of RPPI data:

$$Y_t = (l_{t-1} + b_{t-1}) \cdot (1 + \epsilon_t) \tag{3}$$

$$l_t = (l_{t-1} + b_{t-1})(1 + \epsilon_t) \tag{4}$$

$$b_t = \phi b_{t-1} + \beta(l_{t-1} + b_{t-1})\epsilon_t \tag{5}$$

In these formulas, Y_t is Residential Property Prices Indices of United Kingdom. ϵ_t is the error term at time t. l_t is trend component at time t. b_t is seasonality component at time t. ϕ is damping parameter. β is smoothing parameter for the trend component.

4. Result

This study utilizes rolling forecast to systematically evaluate which model can predict RPPI of United Kingdom more accurately. This rolling forecast approach begins in fourth quarter of 1990 and progressively expands the training set quarter-by-quarter, estimating each model at each step to generate forecasts for the next 5 years. The fitting results of ARIMA, ARIMA with xreg and ETS in the rolling forecast is shown below in Fig. 2, Fig. 3 and Fig. 4, respectively.

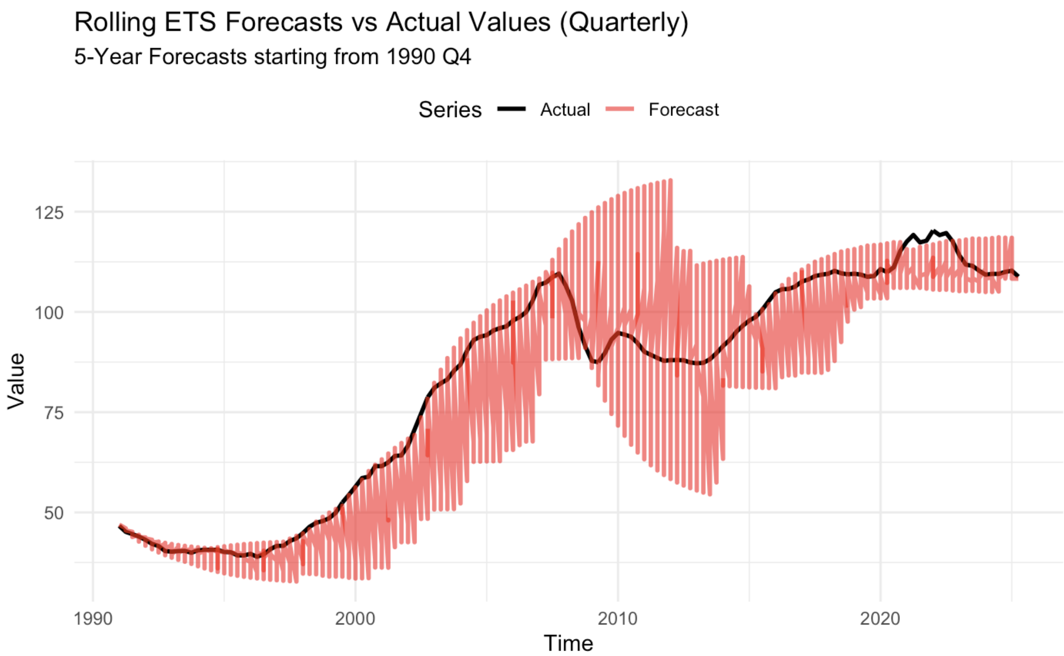

Fig. 2 ETS Result

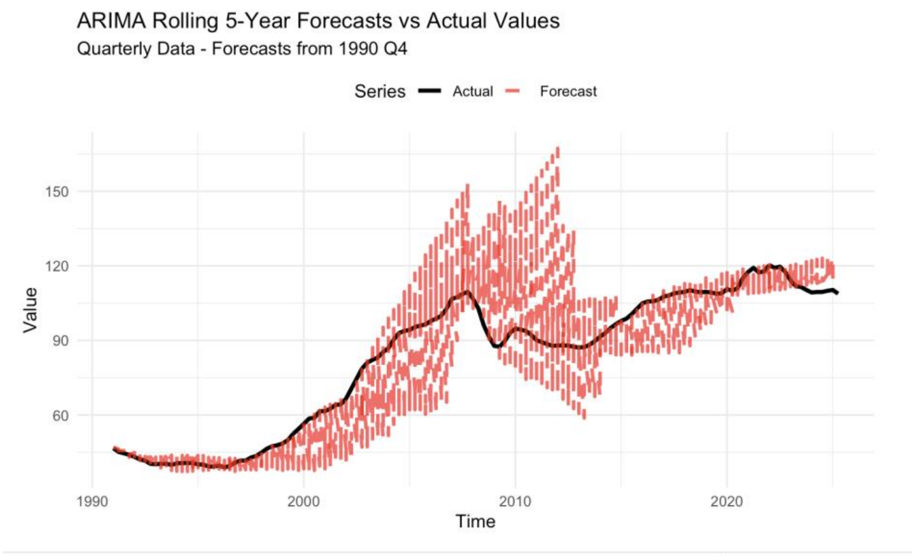

Fig. 3 ARIMA Result

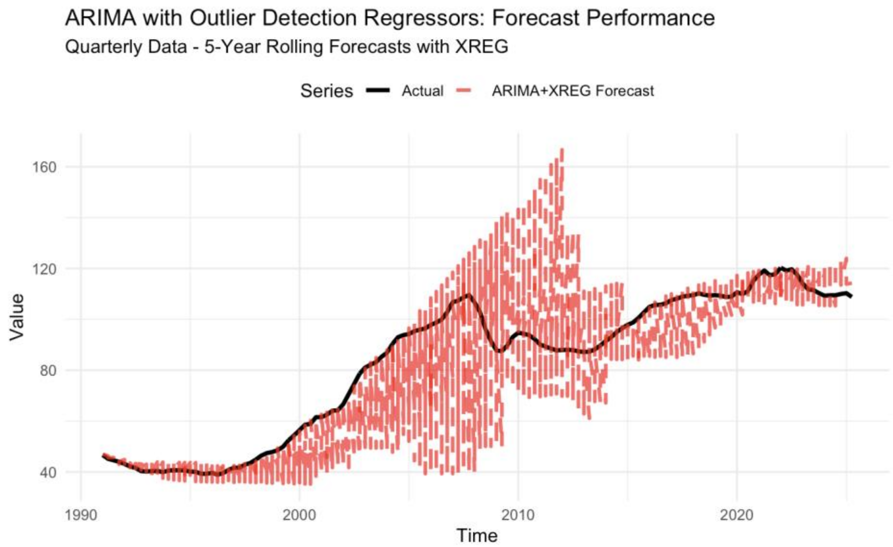

Fig. 4 ARIMA with xreg Result

The three figures above represents that the rolling forecast results of these three models are all capable of tracking the long-term upward trend and major fluctuations of the time series data of RPPI of United Kingdom, which is the same as the actual situation. This paper compares the values of overall RMSE and overall, MAE of the rolling forecast of these three models to further compare the forecasting accuracy of the three models. Overall RMSE is the square root of the average of squared errors from every individual forecast point across all rolling windows, measuring the overall magnitude and dispersion of forecast errors. Overall, MAE is the average of absolute errors from every individual forecast point across all rolling windows, which represents the mean absolute size of forecast errors. The values of overall RMSE and overall, MAE are shown in Table 2.

Table 2. Comparison Results

Model	Overall RMSE Results	Overall MAE Results
ARIMA	14.16469	9.388311
ETS	5.354347	3.03682
ARIMA with xreg	16.11669	10.46971

ETS has both lower overall RMSE and lower overall MAE values, which means ETS can more accurately capture underlying trends, seasonality, and fluctuations and more accurately forecast the time series of RPPI of United Kingdom.

5. Conclusion

Accurately predicting the RPPI can assist England policymakers to develop effective policies. This study employs ARIMA, ARIMA with external regressors (ARIMA with xreg), and ETS for the sake of figuring out the more accurate model for forecasting quarterly data of RPPIs for United Kingdom. The ETS model demonstrates optimal performance, achieving the lowest overall Mean Absolute Error and overall Root Mean Square Error among all evaluated models. But this study does not consider the impacts of other external variables on RPPI forecasting. For instance, macroeconomic variables such as the unemployment rate could be incorporated into the ARIMA model. This involves modeling the effect of past values of these macroeconomic variables on future RPPI changes, which makes the RPPI forecasting more comprehensive and more rigorous.

References

[1] Zamri M, Rifin M, Amit N. Application of the ARIMA model in house price index in Malaysia. Jurnal Intelek, 2024, 19(2): 184-192.

[2] Bełej M. Housing price forecasting in selected Polish cities during the COVID-19 pandemic. Geomatics and Environmental Engineering, 2021, 15(4): 59-80.

[3] Dhamo Gjika E, Puka L, Zacaj O. Forecasting consumer price index (CPI) using time series models and multi regression models: Albania case study. Proceedings of the 10th International Scientific Conference "Business and Management 2018". 2018: 466-472.

[4] Soy Temür A, Akgün M, Temür G. Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models. Journal of Business Economics and Management, 2019, 20(5): 920-938.

[5] Sun Z. Comparison of trend forecast using ARIMA and ETS models for S&P500 close price. Proceedings of the 2020 the 4th International Conference on E-Business and Internet. 2020: 57-60.

[6] OECD. Analytical house prices indicators. (2025) [2025-09-28]. https://data-explorer.oecd.org/vis?lc=en&tm=housing%20price&pg=0&snb=65&isAvailabilityDisabled=false&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_AN_HOUSE_PRICES%40DF_HOUSE_PRICES&df[ag]=OECD.ECO.MPD&df[vs]=1.0&dq=.Q.RHP.&lom=LASTNPERIODS&lo=20&to[TIME_PERIOD]=false.

[7] East S, Zahed M. Forecasting U.S. unemployment rates using ARIMA: A time series analysis from 1948 to 2019. SESUG Paper 2024.

[8] Siami-Namini S, Tavakoli N, Namin A S. A comparison of ARIMA and LSTM in forecasting time series. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). 2018: 1394-1401.

[9] Bógalo J, Llada M, Poncela P, et al. Seasonality in COVID-19 times. Economics Letters, 2022, 211: 110206.

[10] Oh J, Seong B. Forecasting with a combined model of ETS and ARIMA. Communications for Statistical Applications and Methods, 2024, 31: 143-154.