Paris House Rental Price Index Prediction-A Classical Statistical Model Approach

Xiaopeng Zhang *

Student in undergraduate school of mathematics, Sorbonne Université, Paris, France

* Corresponding Author Email: xiaopeng.zhang@etu.sorbonne-universite.fr

Abstract. The study focuses on predicting rental prices in Paris and aims to contribute to urban economics and data analytics. It analyzes a wide range of data sources, including historical rental prices, economic indicators, demographics, and regulations. The goal is to compare classical statistical models' prediction accuracy of these three models: ARIMA, dynamic regression, and random forest. The results reveal that the ARIMA model performs best, offering more accurate predictions. ARIMA relies on time series analysis, capturing complex patterns in rental prices, making it well-suited for dynamic real estate markets. The study also examines the impact of external factors like rent reference indices, house price indices, and unemployment rates on prediction accuracy. While these factors seem promising, the further analysis suggests they can introduce noise into predictions if not chosen carefully. In conclusion, this research contributes valuable insights for rental market stakeholders. The ARIMA model proves effective for rental price forecasting in Paris, emphasizing the importance of understanding intrinsic time series patterns. This study can guide decision-making for renters, landlords, and investors in Paris and similar urban areas, providing a better understanding of predictive modeling in housing markets.

Keywords: Statistical model; data analysis; comparative analysis; housing accessibility.

1. Introduction

The real estate market plays a pivotal role in urban economies, reflecting the intricate interplay of economic, social, and environmental factors [1]. Among the various elements of the real estate market, rental prices hold a significant position due to their direct impact on affordability, housing accessibility, and overall urban planning. As urbanization accelerates globally, metropolitan cities like Paris experience increased demand for housing, putting immense pressure on the rental market [2]. Accurate forecasting of rental prices becomes crucial not only for tenants and landlords but also for policymakers, investors, and researchers aiming to comprehend the dynamics of housing markets.

The City of Light, Paris, renowned for its rich cultural heritage and economic prominence, faces unique challenges in maintaining a balanced housing market. The diverse population, international influence, and changing economic landscapes contribute to the complexity of the city's real estate dynamics [3]. Over the years, advancements in data availability, computational techniques, and predictive modeling have provided new avenues for understanding and predicting rental price trends. This paper delves into the intricate task of forecasting rental prices in Paris, aiming to contribute to the broader discourse on urban economics, data analytics, and housing market dynamics [4, 5].

This study embarks on a quantitative analysis of the Parisian rental market, utilizing a wide range of data sources, including historical rental prices, macroeconomic indicators, demographic shifts, and data that incorporates governmental regulations. By employing multiple advanced predictive modeling techniques, the project endeavors to study the prediction accuracy of different classical statistic models [6].

In addition to the primary objective of comparing classical statistical models in the context of the Parisian rental market, this research yields valuable insights for various stakeholders [7, 8]. The findings can serve as a reference point for researchers and practitioners seeking to understand the performance differences among these models in predicting rental prices. This comparative analysis can guide future model selection in similar urban contexts, leading to more accurate forecasts. Moreover, this study underscores the importance of robust statistical methodologies in real estate

analysis, emphasizing the need for continued exploration and refinement of predictive modeling techniques. As such, this research contributes not only to academic knowledge but also equips professionals with the tools to make data-informed decisions in housing market analyses.

2. Methodology

2.1. Description of Data

All the historical data including French rent index [4], French rent reference index [5], French house price index [6] and France metropolitan unemployment rate [7]. has been collected from INSEE website (Institut National de la Statistique et des Études Économiques). The dataset exhibits a quarterly frequency, spanning approximatively 2 decades, from 4th quarter 2002 to 2nd quarter 2023. The computation of the French rent index is derived from data obtained through the French Rents and Charges Survey and the survey on rents within the social sector (ELBS). It is noteworthy that while the actual numerical value of the French rent index does not directly correlate with prevailing house rental prices, its fluctuations hold substantial significance as an indicator of the dynamics within the house rental market pricing [9]. Conversely, the French rental reference index serves as an indicator that mirrors the governmental regulations pertaining to rent escalation, thereby ensuring that the rental pricing remains commensurate with the evolving purchasing power [10]. Table 1 shows all the data that has been collected and their time period.

Table 1. Data description

Dataset	Time period	Range of values
French rent index	2002-Q4 – 2023-Q1	73.8 – 104.4
French rent reference index		105.61 – 138.61
French house price index		59.9 – 133.9
France metropolitan unemployment rate		7.1% – 10.5%

2.2. Data Preprocessing

Data preprocessing part is essential to any time series analysis as it established the groundwork for meaningful and reliable results. At the beginning of this research, a preliminary imperative was to align the dataset to ensure consistency in preparation for subsequent model building part. Furthermore, due to the study's specific focus on the Parisian house rental market, a deliberate process was undertaken to subdivide the comprehensive French rent index data. This segmentation was executed to isolate and differentiate the Paris rent index from its provincial counterparts, refining the dataset to align with the specific geographic scope under examination.

Another important data preprocessing procedure of notable significance is the train-test partitioning. Within the context of this research investigation, a partitioning ratio of 7:3 is employed on the dataset. This partitioning facilitates an evaluation of the prediction accuracy when employing diverse modeling techniques.

2.3. Prediction Methods

In this study, three modeling techniques are used to have a comparative study on the statistical models to forecast the future values of Paris rent indices.

2.3.1 ARIMA model

The Autoregressive Integrated Moving Average (ARIMA) model is a powerful and widely used time series forecasting technique. It is designed to capture and model complex temporal patterns in data [8]. The ARIMA model consists of three components which are autoregressive part, moving average part and differencing order. If we assume that y_t' is the differenced series with order d, and p, q are respectively autoregressive order and moving average order then the full model $ARIMA(p,d,q)$ can be written as eq.1.

$$y_t' = c + \phi_1 y_{t-1}' + \cdots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (1)$$

2.3.2 Dynamic regression

The dynamic regression method is an important deviation of ARIMA model, as it takes the effects of historical data into account, it can also consider the effects of external regressors that might be relevant. In this study, the rent index is assumed to have a linear relation with rent reference index, house price index and unemployment rate with an ARIMA error term. If the ARIMA error term is noted by η_t, dependent variable and independent variable are respectively noted by y_t and $x_t^1, \ldots, x_t^n$. Then a dynamic regression model can be written as eq.2.

$$y_t = \beta_0 + \beta_1 x_t^1 + \cdots + \beta_n x_t^n + \eta_t \quad (2)$$

3. Results and Discussion

To evaluate the effectiveness of the models, both the predicted values and fitted values are subjected to 4 metrics to evaluate its error compared to train set and test set values. The 4 metrics are respectively Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Bias Error (MBE).

If the error is denoted by ϵ_i, the actual value is denoted by y_t, and the total number of values examined s n. RMSE is computed using eq. 3.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \epsilon_i^2}{n}} \quad (3)$$

MAE is computed using eq. 4.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\epsilon_i| \quad (4)$$

MAPE is computed using eq. 5.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\epsilon_i|}{y_t} * 100 \quad (5)$$

MBE is computed using eq. 6.

$$MBE = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \quad (6)$$

Fig.1 represents the residual diagnosis of the fitted values of model ARIMA (0,2,1). By direct observation, the fluctuations of residual time plot have a shape of white noise, and none of the ACF values breaches the critical levels, in addition the histogram of the distribution of residual values is approximate to the graph of a normal distribution. According to these observations, it is logic to admit that the residual of ARIMA (0,2,1) model is a white noise and the model is well fitted.

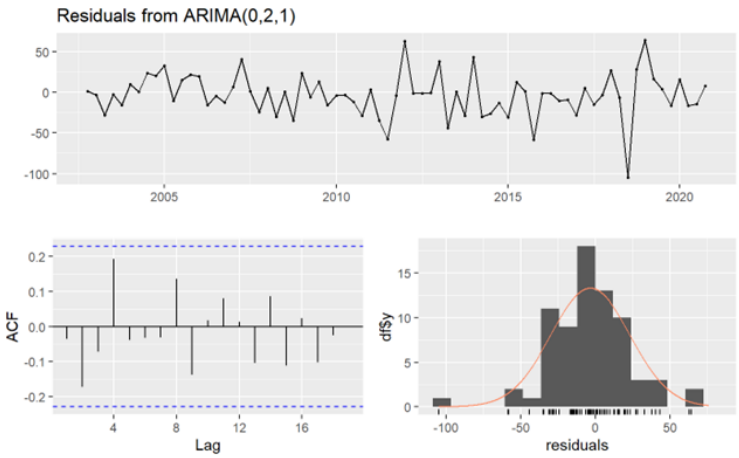


Fig. 1 Residual diagnosis of ARIMA

Fig.2 represents graphs showing the original evolution of Paris rent indices divided into training set (black) and test set (red) and the fitted value using the ARIMA (1,0,1) model along with the forecast (blue) using the model and its confidence interval. The forecast can also be found in blue with its confidence interval.

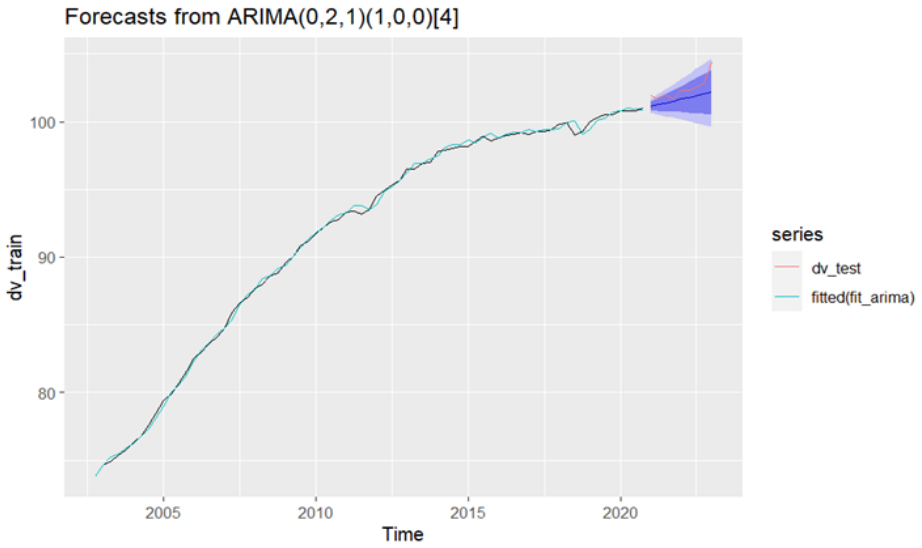


Fig. 2 ARIMA model

Fig.3 represents the residual diagnosis of the fitted values of model ARIMA (1,0,1). By direct observation, the fluctuations of residual time plot have a shape of white noise, and none of the ACF values breaches the critical levels, in addition the histogram of the distribution of residual values is approximate to the graph of a normal distribution. According to these observations, it is logic to admit that the residual of ARIMA (1,0,1) model is a white noise and the model is well fitted.

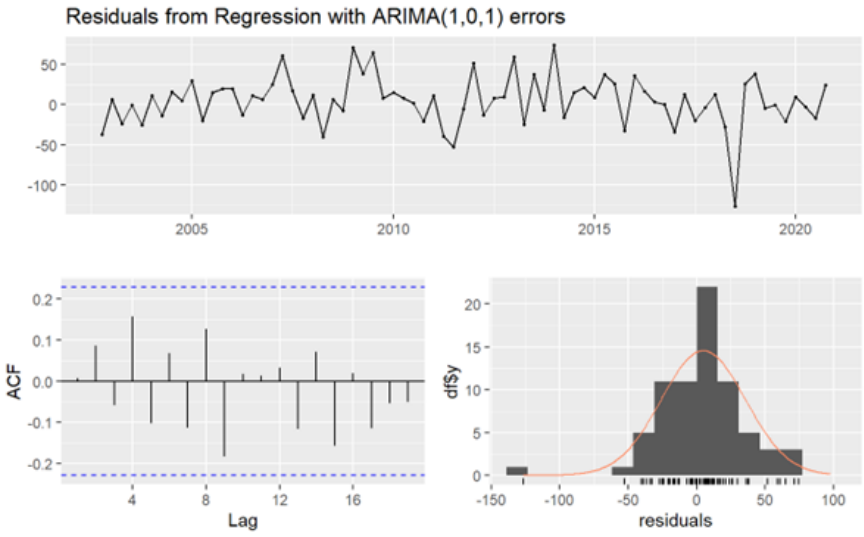


Fig. 3 Residual diagnosis of regression with ARIMA (1,0,1) error

Fig.4 represents graphs also showing the original evolution of Paris rent indices divided into training set (black) and test set (red) and the fitted value calculated using linear regression model admitting Paris rent reference index (IRL), France house price index, and unemployment rate as regressors with ARIMA (1,0,1) model as error term. The forecast can also be found in blue with its confidence interval.



Fig. 4 Dynamic Regression model

A comparative analysis of the metrics which are RMSE, MAE, MAPE, and MBE that measure forecast errors of models is shown in Table 2, it is noticeable that the ARIMA model performed better than the Dynamic regression model in this context.

Table. 2 Comparative analysis of RMSE, MAE, MAPE and MBE obtained.

Model	Dataset	RMSE	MAE	MAPE	MBE
ARIMA	train	0.282	0.204	0.223	-0.033
	test	0.921	0.740	0.719	0.740
Dynamic Regression	train	0.334	0.250	0.275	0.052
	test	1.945	1.584	1.542	-1.417

4. Conclusion

In conclusion, the analysis of the Paris house rental price index has provided valuable insights into the effectiveness of different predictive models, specifically the ARIMA model, dynamic regression, and the influence of external regressors on prediction accuracy. Through a rigorous examination of historical rental price data and various modeling techniques, as it is shown by the result and discussion part, it has become evident that the ARIMA model outperforms dynamic regression in terms of predictive accuracy.

The ARIMA model, which relies on time series analysis and autocorrelation patterns, offers a robust framework for forecasting rental prices in Paris. Its ability to capture temporal dependencies and fluctuations within the data has proven to be a significant advantage, especially in a dynamic market like real estate. This model has consistently demonstrated its superior predictive power, allowing for more accurate rental price forecasts.

Furthermore, the inclusion of external regressors in the modeling process has shown that careful selection is crucial to avoid interference with the prediction results. While external variables such as France house reference index, France house price index, or economic indicators like unemployment rate may initially seem promising for improving predictions, our analysis suggests that they can introduce noise and diminish the predictive capabilities of the model. In some cases, these variables may not exhibit strong relationships with rental prices, leading to misleading conclusions and suboptimal predictions.

Therefore, it is essential to exercise caution when incorporating external regressors into predictive models, ensuring that they are chosen based on rigorous analysis and a solid understanding of their relevance to the rental market. In many instances, the ARIMA model, with its focus on intrinsic time

series patterns, may provide more reliable and accurate predictions than models that incorporate external variables without a well-founded rationale.

In summary, the investigation into the Paris house rental price index highlights the effectiveness of the ARIMA model as a powerful tool for forecasting rental prices. While external regressors can be valuable when appropriately selected, our findings underscore the importance of prioritizing intrinsic time series analysis when aiming for the most accurate predictions in dynamic markets like real estate. This research contributes to a better understanding of predictive modeling in the context of housing markets and offers insights that can inform decision-making for both renters and landlords in Paris and similar urban areas.

References

[1] Charles Ka, Yui Leung. Introduction to the Handbook of Real Estate and Macroeconomics. Handbook of Real Estate and Macroeconomics, 2015, 10-14.

[2] Albis H, et al. International Migration and Regional Housing Markets: Evidence from France. International Regional Science Review, 2019, 42(2): 147–180.

[3] Li Dongyue. Comparative Study of House Price Prediction Models. Industrial Technology and Economics, 2006, 25(9): 3.

[4] Gu Xiujuan, Li Chao. Research on House Price Prediction Based on Markov Chain. Consumer Economy, 2012, (5): 4.

[5] Guo Peijun, Mao Haizhou. Wenzhou Housing Price Prediction Model Based on Grey Theory. Mathematical Practice and Understanding, 2011, 41(20): 7.

[6] Li Huajie, Wu Jiling. Prediction of housing prices in Xiamen based on BP neural network. Journal of Shanxi Radio and Television University, 2011(1): 3.

[7] Gu Ying, Xia Letian. Application of Weighted Markov Chains in House Price Prediction. Journal of Chongqing University of Technology: Natural Science Edition, 2013.

[8] Gu Ying, et al. Application of Weighted Markov Chains in House Price Prediction. Journal of Chongqing University of Technology (Natural Science), 2013.

[9] Li Hengke, Yuan Yuan, Liu Jing, Zheng Yueyan. Prediction and analysis of housing prices in Wuhan based on multiple linear regression models. Science and Public, 2021, 279-280.

[10] Wang Zhichao. Analysis of House Price Prediction Based on Data Mining. Computer Knowledge and Technology, 2017.