Predictive Analytics and Macroeconomic Influence: A Detailed Exploration of the Toronto Housing Market Dynamics

Beiming Yu[1, *], Shenyu Zhou[2]

[1]Department of mathematics, New College, University of Toronto, Toronto, M5S 1A1, Canada

[2]Department of Statistics, University of Florida, Gainesville, 32602, United States

*Corresponding author: beiming.yu@mail.utoronto.ca

Abstract. The housing market, serving as a pivotal entity in economic matrices, inherently exhibits a particular complexity. This research sets foot into a comprehensive investigation of the Toronto housing market, unraveling its intertwining associations with macroeconomic variables while attempting to predict future trends based on such exploration. Backed by a robust data set from January 2011 to July 2023 of the Canadian economy, this study employs correlation models (Spearman and Pearson) and ARIMA to generalize and perform housing price predictions. Preliminary findings via correlation analysis signal a substantial linkage between housing prices and macroeconomic indexes of GDP, employment, and exchange rates. The ARIMA model application, underscored by a p-value of approximately 0.042 from the Ljung-Box test, provides a valid, future price estimation. In the intrinsic puzzle of the housing market, this research offers an alternative understanding from a macroscopic lens. This research, while showcasing predictive prowess, also stands as a testament to the multifaceted nature of the housing market and advocates for the ongoing refinement and diversification of predictive models in navigating its complexities.

Keywords: Housing market; prediction model; macroeconomic factor.

1. Introduction

Toronto, as Canada's largest city and a global hub, has been through remarkable changes in its housing market over the past few decades. Under the trends of urbanization and economic growth, the housing market is becoming volatile and complex, influenced by multidimensional direct and indirect factors [1]. A 2007 research study comprehensively analyzed income polarization in Toronto from the 1970s to the 2000s. This study highlighted the growing gap in housing prices across various neighborhoods, attributing it to significant household income disparities [2]. Due to globalization, heightened international capital flows, and issues inherent to Canada's taxation system [3], the price of Toronto's housing market has become unrelated to local incomes and shifted to investor-driven, leading to the deterioration of housing affordability [3, 4]. More recently, the COVID-19 pandemic intensified the Greater Toronto Area (GTA) housing affordability crisis, worsening the pre-existing problem [5].

For the significance of the housing market to a nation's economy and the world, intensive research generalizing or predicting housing prices was done in the past with models including diversified parameters and determinants. Traditionally, house valuation automation is done with regression-based methods that investigate regular quantified parameters like peripheral facilities, the number of bedrooms, sizes, etc. [6]. Or with standard econometric models that examine macroeconomic indexes' impact on the market [7]. The hedonic pricing model has been fundamental and widespread in inquiring into the factors influencing housing values. Research uses such models to understand the relationship between transportation systems and property prices. One earlier study suggests that the saving in commuting costs is directly reflected in housing values [8]. Another project in 2018 indicates that the average income of the community, its distance to the central business district, and population and employment density factor into the determination of price while pointing out distinctions and heterogeneities in housing stock between houses, townhouses, and condos that different models should be applied respectively [9]. Both studies employed data from the GTA.

Machine learning has been widely adopted in house market predictions. One study in 2018 designed a machine learning system that optimally utilizes linear, forest, and boosted regression models and incorporated the usage of neural networks in further enhancing the algorithm developed [10]. More recently, in 2023, Peng et al. presented LUCE, designed to mitigate the challenges with limited spatial and temporal data of house transactions. The innovative model employs graph convolutional networks (GCN) and long short-term memory (LSTM) networks to extract information from heterogeneous information networks (HIN) and develops a lifelong learning framework for house valuation. Such manipulation made LUCE so comprehensive and competent that it outperformed prior automated methods [11]. Demonstrating the potential and grand application of machine learning in the housing market, these works are examples of expanding the research techniques in real estate following advances in datasets and computation.

The housing market has been heavily influenced by various intrinsic factors that can be intuitive or concealing, and even social media sentiments can influence and predict housing market behaviors [12]. The market in Toronto is a paradigm of such complexity. Detailed in this research, a comprehensive investigation with holistic consideration of impactful factors and employing multiple models for comparative analysis will be performed. Through rigorous and thorough evaluations, this paper attempts to discover the optimal model for predictive accuracy based on data in the Toronto housing market and seek further generalization.

2. Methods

2.1. Data Selection

The data set used in the literature is collected from different websites. The monthly new house price index, monthly gross domestic product index, and monthly unemployment rate come from Statistics Canada. The monthly exchange rate between USD and CAD is collected from the Investing website. The reference period of all the statistics is from January 2011 to July 2023, so the paper uses the data of 150 months.

2.2. Research Protocol

The literature will investigate the relationship between Canadian house price, the dependent variable(Y), and exchange rate, unemployment rate, and GDP, which are the independent variables (X1, X2, X3). Both R code and SPSS will be used on the paper. Based on the R code, this research will use the Spearman or Pearson correlation model to study the correlation between dependent variables and each independent variable. The following content will check whether the variables have linear relationships since the Spearman model measures the monotone relationship between non-linear and continuous variables. In contrast, the Pearson model better estimates the correlation between variables with linear relationships.

After determining the existence of a correlation between the variables, the research will do factor analysis with SPSS to find out the different influences of the variables and obtain a regression to forecast the future housing price. The ARIMA model will also be used to predict the future price, and the paper will compare the results from both methods.

3. Results and Discussion

Table 1 shows the correlation between housing prices and the other three variables using the Pearson and Spearman methods.

Table 1. Correlation between house prices and independent variables

Methods/Variables	Unemployment rate	GDP	Exchange rate
Pearson	-0.332	0.886	-0.863
Spearman	-0.453	0.923	-0.676

The research used the R code to draw the linear trend between house prices and the other three variables to determine the linearity of each pair of variables. And the result found that the p-values of three trends are smaller than 0.05, as shown in Table 2.

Table 2. Linear trends between house prices and independent variables

	Intercept estimate	Standard error	p-value
Unemployment	119.533	3.992	2e-16
GDP	-43.61	63.07	1.32e^-10
Exchange rate	188.915	4.128	2e-16

Thus, the linear trends are statistically reasonable estimations of the relationship between Canadian housing prices and each variable. This paper will choose the Pearson correlation method since it better estimates two variables with a linear relationship. Based on the data in the first row, Table 1, a strong positive relationship between house prices and GDP and a negative one between unemployment and exchange rates can be observed.

The second step is trying to construct the model using SPSS. The result of 4 is shown in Table 3, and the R^2 for the model is 0.939, which implies the model has a high fitting degree with our data. However, the p-value for the coefficient of GDP and month are significantly higher than 0.05. Thus, these two variables should not be included in the model since the value of GDP and month may be independent concerning the house price. After eliminating the GDP index and month index in model 2, a new linear model with an R^2 of 0.936 is developed with the remaining variables; thus, getting rid of the GDP index and month index in the model does not affect the fitting degree. Therefore, the model for estimating the housing price index is:

$$Housing\ price\ index\ =\ -3063.144 - 1.461X1 - 32.647X2\ +\ 1.687X3 \quad (1)$$

Table 3. Model outcomes

Model 1	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-5055.009	772.192	-	-6.546	.000
Unemployment	-2.355	.373	-.326	-6.316	.000
Exchange rate	-29.380	3.092	-.362	-9.502	.000
GDP	-3.372E-5	.000	-.383	-2.676	.008
YEAR, not periodic	2.767	.417	.929	6.634	.000
MONTH, period 12	.017	.068	.005	.245	.806
Model 2	B	Std. Error	Beta		
(Constant)	-3063.144	208.823		-14.669	.000
Unemployment	-1.461	.159	-.202	-9.164	.000
Exchange rate	-32.647	2.879	-.402	-11.340	.000
YEAR, not periodic	1.687	.108	.566	15.639	.000
a. Dependent Variable: house price					

The last work the analysis was trying to estimate the future price index using the ARIMA model with R code. Since the PACF plot shows, the lag-1 autocorrelation is positive and strong positive autocorrelation in the ACF plot. The data was also found that after two times different methods, it will become stable. Therefore, using the ARIMA(0,2,1) to forecast the future housing price index is a better choice compared to other ARIMA models. The result is shown in Figure 1, and based on the Ljung-Box test, the p-value of the forecasting is 0.04172, which is smaller than 0.05. Thus, the forecasting is statistically reasonable.

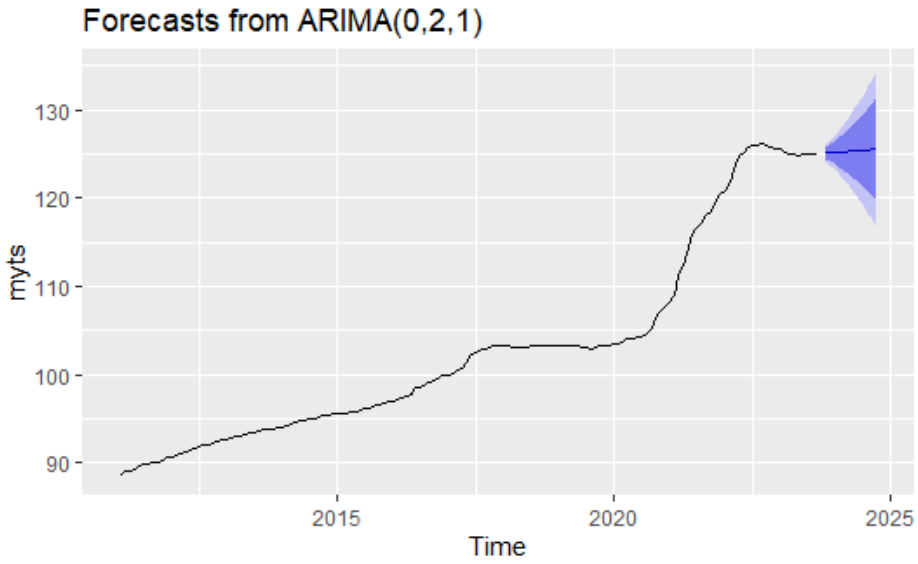

Fig. 1 Forecast graph of ARIMA model

4. Conclusion

The intricate relationship between macroeconomic indicators and the housing market paints a compelling narrative of interconnected dynamics and dependencies. An in-depth exploration of data spanning decades into the Toronto housing market reveals the profound influence of macroeconomic indices, such as GDP, exchange rate, and unemployment rate, on housing prices.

Though meaningful conclusions are drawn herein, it is crucial to acknowledge the limitations embedded within the data and models applied. The data does not exhaustively encapsulate all potential variables and nuances, like other economic parameters that may have distant relationships with Toronto housing prices, and the models fail to factor in specific characteristics of each house, which the demographic and temporal range might introduce biases or overlook specific emerging trends and patterns in the complex market. However, the results of this study should remain valuable. The detailed correlation analysis and an ARIMA model for future price prediction shed light on the association between economic indicators and housing prices, where the evident linear correlations and model-fitting outcomes further underscore this. This study provides insights regarding considering less-explored variables in the context of housing prices. Investigating the macroeconomic trends offers a broad lens through which housing market oscillations can be interpreted.

While the investigation has surfaced several correlations and produced models with high fitting degrees, caution should be exercised in interpreting these results as definitive or broadly applicable because the specific context and potential unseen variables are not accounted for in this study. Despite its strengths, the model might only account for some of the intricacies of the actual market and should be viewed as one of many tools in understanding and predicting housing prices.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Forrest R. Globalization, governance, and the housing market. Governance, Globalization, and Public Policy, 2008.

[2] Hulchanski J D. The three cities within Toronto. Toronto: Cities Centre, 2010.

[3] Gordon J C. Reconnecting the housing market to the labor market: Foreign ownership and housing affordability in Urban Canada. Canadian Public Policy, 2020, 46(1): 1-22.

[4] Gordon J C. Solving puzzles in the Canadian housing market: foreign ownership and de-coupling in Toronto and Vancouver, Housing Studies, 2022, 37: 1250–1273.

[5] Brail S, Kleinman M. Impacts and implications for the post-COVID city: the case of Toronto. Cambridge Journal of Regions, Economy, and Society, 2022, 15: 495–513.

[6] Basu S, Thibodeau T G. Analysis of Spatial Autocorrelation in House Prices. The Journal of Real Estate Finance and Economics, 1998, 17: 61–85.

[7] Gonzalez A. The resilience of microfinance institutions to national macroeconomic events: An econometric analysis of MFI asset quality, 2007.

[8] Bajic V. The Effects of a New Subway Line on Housing Prices in Metropolitan Toronto. Urban Studies, 1983, 20(2): 147–158.

[9] Hawkins J, Habib K N. Spatio-Temporal Hedonic Price Model to Investigate the Dynamics of Housing Prices in Contexts of Urban Form and Transportation Services in Toronto. Transportation Research Record, 2018, 2672(6): 21–30.

[10] Varma A, et al. House price prediction using Machine Learning and Neural Networks. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018.

[11] Peng H, et al. Lifelong Property Price Prediction: A Case Study for the Toronto Real Estate Market. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(3): 2765–2780.

[12] Durai S A. Resale HDB price prediction considering covid-19 through sentiment analysis. European Conference on Social Media, 2023, 10(1): 276-285.