Understanding Housing Bubble in Vancouver, BC: A Time Series Analysis

Zifan Yu[1] and Lily Zhang[2]

[1]Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, MN, USA
[2]Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, MN, USA

Abstract

We characterize housing price trends in Vancouver between 2005 and 2021 using the autoregressive integrated moving average (ARIMA) model, and use the vector autoregressive (VAR) model to explore possible contributions of social and economic indicators. During the 17-year period studied, housing prices in Vancouver have surged. Our results suggest that fundamental factors (of consumer demand) may not be sufficient to account for the trend. We find that purchases based on investors' expectations possibly account for the rise. A real estate transaction tax imposed on foreign buyers in Vancouver in 2016 dampened housing price increases in the short term, substantiating the importance of foreign demand in the market and its effect on price trends.

Keywords: Vancouver house prices; market expectation; Bill 28; time series; ARIMA; VAR; Keynesian economics

"History Doesn't Repeat Itself, but It Often Rhymes"

– Mark Twain.

1. Introduction

Housing affordability is a growing problem in many cities around the world. Vancouver's rapidly inflating housing prices are an extreme case; the most extreme in North American residential real estate. In 2016, when the ratio of single family home price to mean household income reached 13 to 1, Vancouver's housing became the third most unaffordable in the English-speaking world (Todd 2017). A poll in June 2015 by Angus Reid showed that 75% of resident respondents agreed with the statement that real estate prices were unreasonably high (Gordon 2016). Based on the latest statistics, the benchmark price for a typical Vancouver home was $379,900 in 2005, $648,500 in 2015, and reached $1,345,700 in March 2022.

Nevertheless, identifying the drivers behind the staggering price increases can be complicated.

On one hand, the argument that foreign money has been pushing up prices in Vancouver and Toronto has been around for years, and has become a hot political issue.

On August 2, 2016, the government of British Columbia announced and imposed a 15% foreign buyer transfer tax in Metro Vancouver to reduce the rise in home prices caused by foreign investment. While this may reduce further home price inflation, it may not reduce or relieve existing inflated housing prices. Hu (2018) presented a fixed effects model and difference-in-differences design using a panel dataset that consists of 1,482 observations from January 1977 to February 2018 from customized tabulations with several macroeconomic indicators as control variables. The study shows the tax had a weak positive causal relationship on total book value of properties and a strong positive effect on the housing supply; the act did make more housing available.

On the other hand, unlike similar situations in other cities, studies on common factors related to house price suggest that supply constraints and standard demand fundamentals do not explain rapid price increases well (Gordon 2020a).

Additionally, according to classical economic theory the price of goods and services in the market depends on demand in the long run. Nevertheless, Keynes, in his work, suggested that people are influenced by sentiments and opinions that can cause volatility in market prices (Keynes 1936). In 1988 and 2003, Karl E. Case and Robert J. Shiller conducted surveys to investigate home buying behaviors, with the results indicating that buyers were influenced by an investment motive and had strong expectation for future appreciation with little risk perceived in uptrending markets, which could cause house market bubbles (Case and Shiller 2003). Their idea may suggest a third possibility in the case: The price relies on its own momentum to a large extent due to investment motives as the uptrending market appeals to higher expectations. Anticipation of continued increases in housing valuations drove investments and valuations.

We believe this puzzle is worthy of empirical analysis. Although most research concerning the possible driving force for house price in the Vancouver real estate market use customized panel data and quasi-experimental methods, these research designs are poorly accessible and subject to endogeneity bias. In this project, we will first analyze the characteristics of price trend internal characteristics with an autoregressive integrated moving average (ARIMA) model. Further, we develop a Vector autoregressive (VAR) model that relates Vancouver house price to the multi-time series of common economic and social factors that are hypothesized to affect house price. The interdependence and relationships are analyzed through the Granger causality test, impulse response, and variance decomposition.

The remainder of the project is formulated as follows: Section 2 states the data sources and the metrics they are based on and methodology specifications. Section 3 demonstrates our modelling, our assumptions, their validity, and results. Section 4 reflects on the strength and weakness. Section 5 concludes with discussion and contemplation.

2. Data and Methodology

Data

This study uses monthly open data from January 2005 to December 2021 that is available from official agencies. In total, we have eight endogenous variables for 204 months; 1,632 observations are included, with no missing data. In addition, we add a dummy variable "policy," which equals 1 from August 2016 onward (and 0 before), when Bill 28 became active, in the VAR model to assess and predict the impact of the policy. We analyze monthly log differences of house prices, $\Delta \log(HPI_t) = \log(HPI_t) - \log(HPI_{t-1})$; multiplying by 100 gives an approximate monthly percent change for interpretation.

House Price Data House price data are downloaded from the Canadian Real Estate Association (CREA) official website. CREA offers public access to the MLS® House Price Index (HPI), which is constructed by defining "typical" homes (benchmark property) based on homes bought and sold of different features through the MLS® System data and sophisticated statistical models. Outliers are excluded. They construct sophisticated models to include quantitative property attributes that include things such as square feet of the home and the most commonly occurring value (i.e. modal value) for qualitative attributes, which could include unfinished portions of the home (REALTORS 2022).

House price in the study refers to the monthly composite benchmark prices in Greater Vancouver, an aggregation of benchmark categories and metropolitan markets tracked by the Index (CREA 2016).

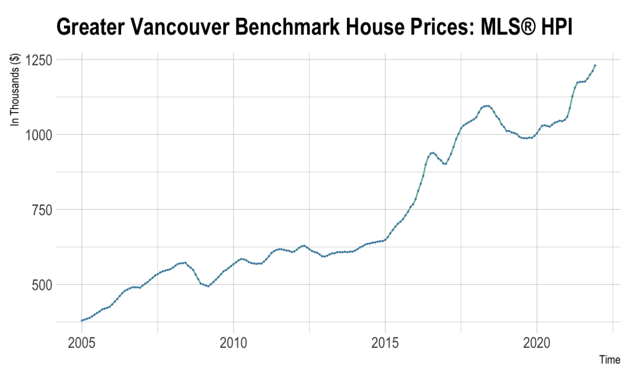

Figure 1 The monthly benchmark house prices in Greater Vancouver from January 1, 2005 to December 1, 2021.

Economic and Social Indicators In the hope of exploring potential drivers behind the house price trend, we consider several economic and social indicators theoretically related to house price from demand fundamentals and social development perspectives.

In specific, we consider

(a) Gross domestic product (GDP) using 2012 as the reference year basis.

Gross domestic product (GDP) is a widely used measure of economic production. It is important to note here that the "gross" in GDP relates in this study to capital consumption costs, which are the costs associated with the devaluation of capital assets such as building materials and machinery (Government of Canada 2022a). The monthly GDP data are available on Statistics Canada, the national statistics agency that provides Canadians with key information on Canada's economy, society, and environment. GDP serves as an important indicator for the general development of the economy and citizens' general ability to consume.

(b) Long-Term Government Bond Yields: 10 Years for Canada.

The return an investor realizes on the 10-year government bond from the Federal Reserve of St. Louis Economic data (FRED). Based on non-arbitrage conditions, it effectively reflects the interest rate or the cost of borrowing.

(c) Canadian Dollars to U.S Dollar Spot Exchange Rate.

The data is available in the daily bilateral exchange rates and U.S. dollar indexes, which is released weekly on every Monday at 4:15 p.m. by the Federal Reserve Board.

The United States is Canada's biggest trading partner, accounting for more than 70% of Canadian international trade volume (FederalReserveBoard 2022). US currency value is strongly correlated with the strength of the Canadian Dollar. In transactions involving foreign investors, e.g. real estate purchases in Vancouver, the strength of the Canadian Dollar is an important factor affecting purchase costs.

(d) Consumer Price Index(CPI) for Canada using 2015 as base year (100).

The Consumer Price Index (CPI) indicates changes in consumer prices, and is formulated by comparing the cost of a fixed basket of goods and services purchased by consumers over time. It is widely used to characterize the change in the general level of consumer prices or of the rate of inflation (Fernando 2022).

(e) Weekly earnings in British Columbia.

Weekly earnings is released by the Survey of Employment, Payrolls and Hours (SEPH), which provides a monthly snapshot of the amount of earnings, and the number of jobs and hours worked by detailed industry at the national, provincial, and territorial levels.

Average weekly earnings are derived by dividing total weekly earnings by the total number of employees. Earnings of employees serve as the direct source for consumption and investment. If the average earning increases, it is expected that some of these incomes will flow to the real estate market and drive the price up (Government of Canada 2022b).

(f) All employment, including overtime, in British Columbia

The employment statistics are released from SEPH.

The calibration is done using a generalized regression estimator. The model groups are mostly defined at the national and sub-sector levels (i.e., three-digit North American Industry Classification System (NAICS) code or, in a few instances, four-digit); in a few cases, the enterprise size (employment) and the provincial dimensions are utilized. Calculating at the model group level, regression coefficients are applied to the estimates of total employment and payrolls from the administrative sources to determine the additional variables (authority of the Minister 2021).

We include the variable for the following considerations: First, the numbers of people employed suggest the number of potential consumers. More plainly stated, an employed person will be more likely to purchase a new house or invest in the real estate market than an unemployed one. Second, the employees of the economy, depicted as the labor force by economists, signal the strength of future development, essentially contributing to the economy's growth (Solow 1956).

(g) Crime cases in Vancouver

The crime data is an aggregate data accessed from VPD Crime Data Vancouver Police Department. Crime data are reported by case at the neighborhood level by types of crime. We manually aggregate cases of all types of crime in each month for all neighborhoods in Vancouver.

Crime is closely associated with house price as follows: First, potential consumers may be less likely to choose Vancouver as their residence if they observe an increasing trend in its crime rate. Second, criminological theory suggests that increasing unemployment rates closely correlate with increasing crime rates and further affect economic growth and consumption (Becker (1968); Freeman (1996)).

Bill 28 (Foreign-buyers Tax) Previous research has identified foreign ownership as an influential factor in the Vancouver housing market. Gordon, J. C. (2020b) conducted regressions on Canadian housing prices with time-series cross-sectional analysis, with the results showing that supply constraints and standard demand fundamentals cannot account for the rapid increase in price, which leaves two other explanations: foreign demand and speculations. The research provides evidence that at least 10% of sales are by foreign buyers in most regions of Vancouver and Toronto in 2016.

On August 2, 2016, the British Columbia provincial government introduced Bill 28, the Miscellaneous Statutes (Housing Priority Initiatives) Amendment Act to impose 15% additional property transfer tax on foreign home buyers in the Greater Vancouver area. In February 2018, the BC government increased the foreign buyers' tax to 20% and expanded the range to Fraser Valley, the Capital and Nanaimo Regional Districts on Vancouver Island, and the Central Okanagan Regional District (Wikipedia contributors 2021).

R Packages In addition, the required R packages for our models are **dplyr**, **ggplot2**, **hrbrthemes**, **astsa**, **lubridate**, **openxlsx**, **timeSeries**, **tseries**, **vars**, and **lmtest**. The **dplyr**, **ggplot2**, and **hrbrthemes** packages are installed for data manipulation, visualization creation, and typography, respectively. Since our data is cleaned in an Excel file, we use the **openxlsx** package for simplifying the process of writing and styling Excel (.xlsx) files. Specifically to our time series analysis, we choose the **lubridate** and **timeSeries** package because they are optimized for working with dates and times. The **astsa** package is used for implementing time series models and conducting time series related statistical tests. The **vars** package is installed for VAR modeling. The **tseries** package is to access the Augmented Dickey-Fuller test for the null hypothesis that x has a unit root. The **lmtest** package is installed for performing a test for Granger causality.

Methodology

To analyze the underlying cause of trends and features of Vancouver house market over time, we adopt time series models that seek to capture the features of the series by applying past observations to explain and characterize the future values. The specifications are explained in detail as following:

Autoregressive Integrated Moving Average (ARIMA) To capture the most variations, we hypothesize an ARIMA model in seasonal context. Specifically, AR stands for the idea in which the target variable is regressed on its own past values, which implies that lagged values are used to make predictions. MA, abbreviation for moving average, means that the lagged prediction errors are put as inputs. 'I' stands for 'integrated' which refers to differencing to remove the trend. Seasonality is handled via seasonal AR terms at lag 12.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \quad (1)$$

Equation (1) demonstrates general mathematical equation for ARMA model, from which ARIMA and seasonal ARIMA (SARIMA) models build on. $\{Y_t\}$ is a mixed autoregressive moving average process of orders p and q.

Vector Autoregression (VAR) Advocated by Sims (1986), Vector Autoregression (VAR) model can be interpreted as a natural extension of the univariate autoregressive model to dynamic multivariate time series. With the VAR model, we can feasibly carry out a dynamic analysis of a system in which changes to a particular variable are related to changes in its lags, other variables, and their lags. (Penn State 2022). As the predication of each time series is based on its own past values, or more precisely put as using its own lagged values to predict, the VAR is within the system of autoregression but involve a numbers of time series in a smaller system. As each time series in the system is put in as endogenuous variables, VAR models are bi-directional in essence.

The estimation of a VAR is an Equation-by-Equation Ordinary Least Squares (OLS) regression. Specifically, each variable is regressed on its lags and lags of other variables in the system.

With these properties, VAR has several advantages over univariate autoregressive models. First, it can be used to explore the relation between the variable of interest and a set of variables and no explicit interpretation needed for them. Then, based on the system constructed by it, we can apply the Granger causality test to examine whether a variable contains information for future values of another; further, we could readily conduct impulse response analysis to assess how and to what extent one variable respond to a sudden but temporary shock in another. Lastly, the importance of variables in explaining the volatility of others could be demonstrated explicitly through variance decomposition. (Zhu 2022). Therefore, VAR can capture the evolutionary history and the inter-dependencies between multiple time series, rendering it efficacious for characterizing the dynamic behavior of economic and financial time series and for the subsequent application (Zivot and Wang 2006).

In addition, as we mainly focus on the macroeconomic and social indicators in the study, that can influence each other. In such cases, OLS analysis may succumb to endogeneity, being influenced by variables within a system, without identifying efficacious instrumental variables. However, as the VAR model assumes that all variables are being influenced by others in the system (Dreger and Wolters 2008), it can effectively avoid the bias caused by endogeneity. Therefore, it is reasonable to believe that VAR can better capture and shed light on the house price trend in our context.

Mathematically, the object of interest in the following is the VAR(p) model,

$$Y_t = \nu + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t, \quad t = 0, \pm 1, \pm 2, \ldots, \quad (2)$$

where $Y_t = (Y_{1t}, \ldots, Y_{Kt})\prime$ is a random vector of size $(K \times 1)$, the A_i are fixed coefficient matrices of size $(K \times K)$, and $\nu = (\nu_1, \ldots, \nu_K)\prime$ is a fixed vector of size $(K \times 1)$ of intercept terms. The last term, $u_t = (u_{1t}, \ldots, u_{Kt})\prime$ is a K-dimensional white noise (Lütkepohl 2005).

3. Modeling

ARIMA Modeling

Augmented Dickey-Fuller (ADF) Test The validity of time series analysis relies on stationarity. The initial data obviously demonstrate non-stationarity. Although graphs provide us an intuitive and effective way to deduce and confirm stationarity, to add rigor to our analysis, we use the Augmented Dickey-Fuller (ADF) test to more formally test for it. It is an efficient instrument for testing numbers in time series. It determines the presence of a unit root, a stochastic trend that makes systematic patterns unpredictable in the data generating mechanism by using the ordinary least squares (OLS) estimator obtained by fitting the regression equation to the observed stretch of data (Paparoditis and Politis 2016).

The null hypothesis of the ADF test is that the time series contains unit roots, and the alternative hypothesis statement is that the time series does not contain unit roots. If the p-value from the test is less than a certain significance threshold, we can reject the null hypothesis that concludes the series is stationary. However, if the p-value is greater than the significance threshold, we then accept the null hypothesis that the series is non-stationary. The intuitive idea is that the current value of a series is equal to its last value plus an error if the data contain a unit root. In other words, it has some time-dependent structure and does not have constant variance over time, which provides no relevant information. Thus, it tells us the data should be regressed on deterministic functions of time or differenced to render the data stationary.

We apply differencing at the first order of the original house price for detrending purposes, stabilizing the mean by removing changes in the level of the series. Seasonality is handled via seasonal AR terms at lag 12; we do not apply seasonal differencing. Then the ADF test results for the differenced house price series suggest we could reject the null hypothesis at 95% level and conclude that the series is stationary. Random errors signaling white noise are also observed on auto-correlation function (ACF) plots, which confirms the conclusion.

Autocorrelation Function (ACF) After we detrend to ensure that the time series we use are stationary, we may dig into their autocorrelation and partial autocorrelation, which enables us to specify parameters for the ARIMA model.

ACF gives us values of autocorrelation of any time series with its lagged values. ACF demonstrates the correlation of a sequence with itself, with consideration of the lag by some number of time units. In other words, it represents how well the present value of the series is associated with its past values (Salvi 2019). PACF stands for partial auto-correlation function. PACF demonstrates the same correlation with all intermediate effects removing the direct effect of one to the other. So if there is any hidden information in the residual that can be modeled by the next lag, we might get a farly strong correlation and we will keep that next lag as a feature while modeling (Lazzeri 2021).

The patterns of ACF and PACF imply certain characteristics of lag order for our model. If the ACF plot declines gradually and the PACF plot drops instantly, it suggests that the lag that just cut off captures the variation so well that we don't need past lags to predict the present (Salvi 2019). Then the Auto Regressive model is used. If the ACF plot drops instantly and the PACF plot declines gradually, then we deduce that the information cannot be captured by direct lag correlation and certain residual components explain the trend effectively. Therefore, the Moving Average model is used. If both ACF and PACF decline gradually, then the combined Auto Regressive and Moving Average models (ARMA) is used.

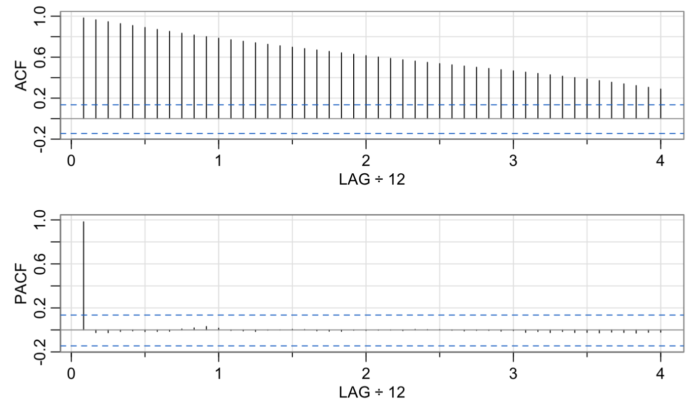

Figure 2 Sample ACFs and PACFs of benchmark housing prices. Lag axes are in terms of years.

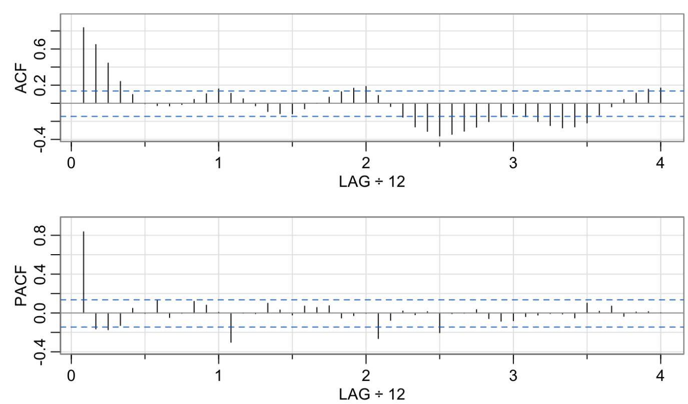

Figure 3 Sample ACFs and PACFs of benchmark housing prices monthly growth rate (after differencing). Lag is in terms of years.

Model Selection The next step is to identify preliminary values of the autoregressive order, p and the moving average order, q. Inspecting the sample ACF and PACF, ACF declines gradually and tail off at 3, PACF drops instantly after lag 1. These results suggest that we fit AR (1), AR(2), and AR(3) models. Our model selection criteria is based on ACF of residuals, p-values from the Ljung-Box test, AIC, and BIC. A good model should produce residuals with insignificant autocorrelation, which could be examined using ACF of residuals and p-values from the Ljung-Box test. Meanwhile, it should produce a low AIC and BIC, which are measures of prediction errors. In the case of conflicting conclusions suggested by AIC and BIC, we chose to favor the model with a lower BIC since BIC gives greater penalization for model complexity. We noticed that the residuals of the ACF plot shows peaks at lag $12s$ ($s = 1.0, 2.0, \ldots$), which suggests to us that it fits a seasonal AR model.

Figure 4 displays a summary plot of the selected model that includes the standardized residuals, the ACF of the residuals, the normal Q-Q plot of the standardized residuals, and the p-values associated with Ljung-Box statistic, at lags $H = 5$ through $H = 36$. First observe from the time plot of the standardized residuals in Fig. 4, in general, there is no particular pattern found. Note that there may be just a few possible outliers, with only a little exceeding 3 standard deviations in magnitude. However, it is optimal that the ACF of the residuals shows no apparent departure from the model's assumptions, except there are 2 slightly higher observations. The straight-line pattern observed from the normal Q-Q plot of the standard residuals suggests that the assumption of normality is reasonable, apart from the few possible outliers. To conclude, based on our model selection criteria, the model appears to fit well.

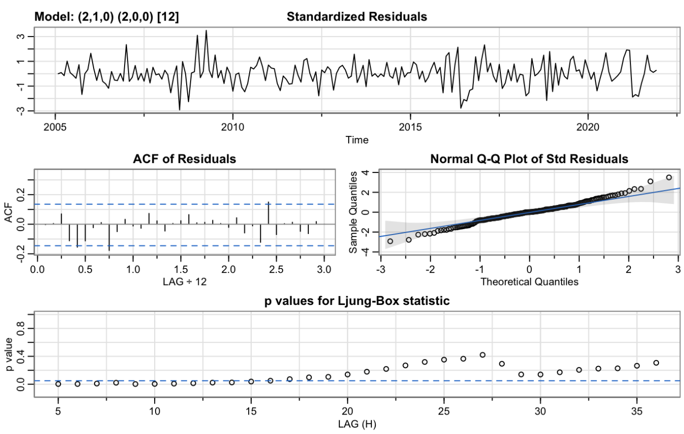

Figure 4 Diagnostics of the residuals from ARIMA $(2, 1, 0) \times (2, 0, 0)_{12}$ fit on benchmark housing prices growth rate.

Forecasting Finally, we forecast the data out twelve months, and the results are shown in Fig. 5. From the forecast plot, we find the Housing Price growth rate is predicted to continue to increase in the next 12 months in 2022.

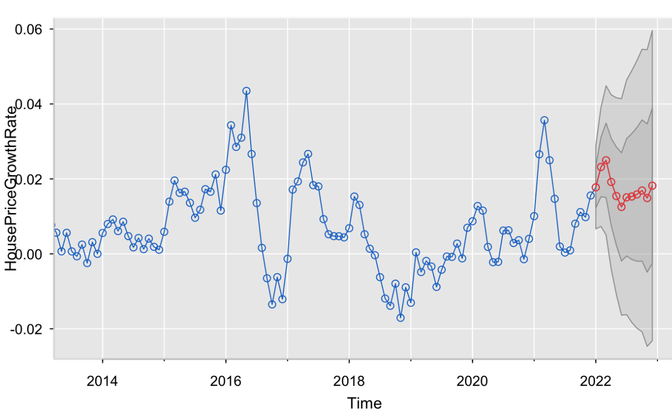

Figure 5 Twelve month forecast using the ARIMA $(2, 1, 0) \times (2, 0, 0)_{12}$ model on benchmark housing prices growth rate.

VAR Modeling

Hypothesized Model

$$\begin{bmatrix} \mathrm{GDP}_t \\ \mathrm{BondYields}_t \\ \mathrm{ExchangeRate}_t \\ \mathrm{Employment}_t \\ \mathrm{Earnings}_t \\ \mathrm{CPI}_t \\ \mathrm{Crime}_t \\ \mathrm{HPI}_t \end{bmatrix} = a_0 + A_1 \begin{bmatrix} \mathrm{GDP}_{t-1} \\ \mathrm{BondYields}_{t-1} \\ \mathrm{ExchangeRate}_{t-1} \\ \mathrm{Employment}_{t-1} \\ \mathrm{Earnings}_{t-1} \\ \mathrm{CPI}_{t-1} \\ \mathrm{Crime}_{t-1} \\ \mathrm{HPI}_{t-1} \end{bmatrix} + \cdots + A_k \begin{bmatrix} \mathrm{GDP}_{t-k} \\ \mathrm{BondYields}_{t-k} \\ \mathrm{ExchangeRate}_{t-k} \\ \mathrm{Employment}_{t-k} \\ \mathrm{Earnings}_{t-k} \\ \mathrm{CPI}_{t-k} \\ \mathrm{Crime}_{t-k} \\ \mathrm{HPI}_{t-k} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \\ \epsilon_{5,t} \\ \epsilon_{6,t} \\ \epsilon_{7,t} \\ \epsilon_{8,t} \end{bmatrix}, \quad (3)$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{20} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{30} & \vdots & 1 & 0 & 0 & 0 & 0 & 0 \\ \beta_{40} & \vdots & \vdots & 1 & 0 & 0 & 0 & 0 \\ \beta_{50} & \vdots & \vdots & \vdots & 1 & 0 & 0 & 0 \\ \beta_{60} & \vdots & \vdots & \vdots & \vdots & 1 & 0 & 0 \\ \beta_{70} & \vdots & \vdots & \vdots & \vdots & \vdots & 1 & 0 \\ \beta_{80} & \dots & \dots & \dots & \dots & \dots & \phi_{80} & 1 \end{bmatrix}.$$

(We identify orthogonalized impulse responses via a Cholesky decomposition consistent with this recursive ordering: earlier-ordered variables can contemporaneously affect later-ordered ones, but not vice versa.)

We first conduct ADF tests on each of the explanatory variables, at the 5% significance level, the ADF statistics of house price, GDP, Employment, Crime, Earnings, BondYields, ExchangeRate, and CPI are all greater than the threshold, indicating that these variables all have a unit root, all variables are non-stationary. We then use first order differencing to remove trends and to produce stationarity. After the first-order difference, the ADF statistical test values of all variables are smaller than the threshold at the same significance level; the variables do not contain unit root after the first difference, indicating that house price, GDP, Employment, Crime, Earnings, BondYields, ExchangeRate, and CPI are all $I(1)$ sequences. Thus, we enter the first difference series into the VAR model. We estimate the VAR in first differences (with a constant) and do not model cointegration, so results should be interpreted as short-run dynamics.

Lag Selection To decide the lag order for the VAR model, we use the VARselect function in R. The function estimates the VAR model of all lag orders smaller than the maximum we assigned (max lag = 8) and automatically compares four information criteria: Akaike information criterion (AIC), Hannan-Quinn information criterion (HQ), Schwarz information criterion (SC) and Final Prediction error (FPE) to give us which lag order minimizes the four criteria, respectively.

In terms of our model, both AIC and FPE suggest optimal lag order of 2, and both HQ and SC suggest optimal lag order of 1. However, if we look closely at the values of AIC criteria, the value of lag 1 is slightly greater than lag 2 (<0.002), which will not significantly impact much on the accuracy of the model. The values of HQ and SC criteria are obviously smaller in lag 1 than lag 2. Therefore, we decided to adopt a lag order of 1 for our VAR model.

The idea of these information criteria rely on the concept of maximum likelihood. We can intuitively understand it based on the ordinary least squares: It tries to assess the suitability of parameters of function/approximation to capture the data patterns. When the noise of our model is in Gaussian distribution, it takes the form of $\ln(p(\mathsf{y} \mid \mathsf{x}, \mathsf{w}, \beta)) = \ln(\Pi_{i=1}^{N} \mathcal{N}(y_i | y(x_i, \mathsf{w}), \sigma^2))$, which is equivalent to OLS estimation. In other words, it estimates how strong our model fits the data.

The more complicated the model, the higher the Maximum likelihood. However, as the complexity of the model can cause overfitting that causes issues when testing data, we need to adjust for parameters. It is what BIC, HQ and SC do based on different metrics.

Granger Causality Usually, regression results only show whether variables are related and how much they are related. As we select appropriate lag order for our VAR model, we then proceed to the Granger Causality test, an important tool to determine their causality in an economic sense. The test was first proposed by 2003 Nobel Economic Prize Winner Sir Clive Granger in 1969.

To illustrate the idea, we may consider two time series X and Y. If the hypothesis test on lagged values of X together with lagged values of Y implies that X contains information about future values of Y on certain significance level, we could suggest that X Granger-cause Y. In other words, when we compare the predictions of Y using its past values with that of Y through its own past values together with lagged values of X, if we find the latter is more accurate than the former, it suffices to say that X Granger-causes Y.

The test implies two preliminaries: the cause take place before the effect; unique information about Y the future values of Y is provided by the cause.

Note that the causality does not necessarily represent exact causation but significance in forecasting.

To clarify, we may compare the conceptual model between the AR model and Granger Causality Test.

$$\text{AR: } y_t = \alpha_0 + \sum_{i}^{p} \beta_i y_{t-i} + \epsilon. \tag{4}$$

$$\text{Granger: } y_t = \alpha_0 + \sum_{i}^{p} \beta_i y_{t-i} + \sum_{j}^{p} \gamma_j x_{t-j} + \epsilon. \tag{5}$$

In this case, the null hypothesis of the Granger test is $H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_p = 0$, and the alternative hypothesis is H_1 : at least one $\gamma_j \neq 0$.

If the p-value from the F-test is smaller than 0.05, we reject the null hypothesis and conclude that at least one $\gamma_j \neq 0$. That means the $\sum_j^p \gamma_j x_{t-j}$ component has some power to use x to explain y and contains information about the future value of y. If we cannot reject the null hypothesis from the F-test ($p > 0.05$), then we cannot say X Granger-causes Y at the 95% significance level.

In our model, we test whether each variable Granger-causes housing prices in the HPI equation of the VAR. Based on the p-values, GDP predicatively causes House Price at 95% significance level; Employment predicatively causes housing prices at 90% significance level. We conclude that only GDP and employment rate are valuable for individually forecasting statistically significant house prices.

Impulse Response Analysis We've seen earlier that Granger causality tests may not tell us the full story about the interactions between (interdependency of) variables. Further, we need another instrument to characterize the dynamic relation in the system in a more sophisticated manner. The interpretation of Granger causality results relies on the assumption that only variables and its lagged counterparts are considered in the analysis, neglecting other potential interdependency from other six variables within the VAR model we constructed.

In response to this limitation, we introduce impulse response analysis that enables us to know the response of a variable to an impulse in another variable within the same model that involves several further variables. Intuitively, impulse response characterizes the reaction of a smaller system (aka part of the model), in response to an external change (the change outside of the small system). It allows us to trace the transmission of a single shock within a system of equations.

Specifically, it identifies the effect on current and future values of the endogenous variable of one standard deviation shock in one of the variables.

Generalized impulse responses (GIRs) are order-invariant; in this paper we use *orthogonal* impulse responses based on a Cholesky decomposition, which are order-dependent. In a lower-triangular Cholesky scheme, earlier-ordered variables can contemporaneously affect later-ordered variables, whereas later-ordered variables do not contemporaneously affect earlier-ordered ones. Therefore, the results of an OIR might be sensitive to the order of the variables.

In our model, we decide the order of variables based on traditional economic theory: for example, we put employment and CPI before crime, since normally lower employment and higher CPI will increase the instability of society so that the crime rate may increase. Nevertheless, the crime rate may not directly cause an increase in CPI or employment in a short period of time.

As shown in the figures 6, 7, and 8, x axis indicates the period after the shock and y axis shows the amount of change due to the shock in units of standard deviations with the red dashed line representing a 90% significance level. The effects of shock from all variables except CPI, bond yields, and GDP on House price are not significantly different from 0.

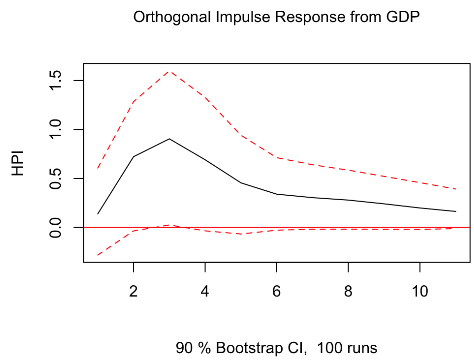

Figure 6 Result of orthogonal impulse response from GDP to house prices (90% bootstrap confidence interval).

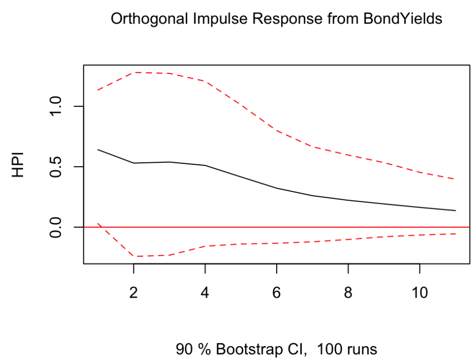

Figure 7 Result of orthogonal impulse response from bond yields to house prices (90% bootstrap confidence interval).

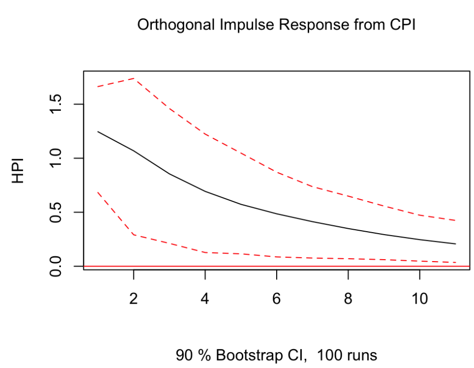

Figure 8 Result of orthogonal impulse response from CPI to house prices (90% bootstrap confidence interval).

Given one standard deviation increase of GDP, house price will respond with about 1 standard deviation increase at the third period with statistical significance but the response becomes insignificant. Similarly, the response to bond yields is also short but instant in period 1; about 0.7 standard deviation. The influence from CPI lasts longer. House prices respond instantly and last 4 periods at 90 percent significance level.

This captures how house prices respond to shocks. For instance, if you see one standard deviation of increase in GDP, if the actual life case is our simulated case, you may expect the strongest increase in the growth rate of house price to occur in period 3. Nevertheless, if the shock is Bond Yields or CPI, you may expect the strongest response instantly and decay shortly.

Also, we observe a positive response to bond yields (interest rate), which seems counter-intuitive. However, it may imply that consumers are not sensitive to the mortgage rate. It can possibly be explained in two ways: First, there are a large proportion of foreign buyers who hardly get a mortgage in Canada; they usually transfer their wealth from their home country to Canada directly and do not need a mortgage. Second, it can be interpreted that the market is a seller's market, meaning that the demand is stronger than the supply, and buyers are not sensitive to price changes.

Variance Decomposition To further capture the relative importance of each endogenous variable in explaining the trend of house price, we present Variance Decomposition for house price forecast of the next 10 period. The forecast error for growth rate of house price is decomposed into components for the contribution of each endogenous variable, through which the proportion of contributions to house price and how they change with time in the autoregressive process is demonstrated. (Eviews 2020).

Figure 9 shows the relative contribution of the variation in benchmark house prices from each shock by decomposing the variance of the forecast error for the growth rate of house prices with respect to house price, crime, CPI, earnings, employment, exchange rates, bond yields and GDP. According to the plot, house price volatility is mostly contributed by its own shocks. Its effect remains for approximately 75% after 10 months. Though other variables show weaker importance, CPI and GDP seem more significant in explaining the housing price volatility

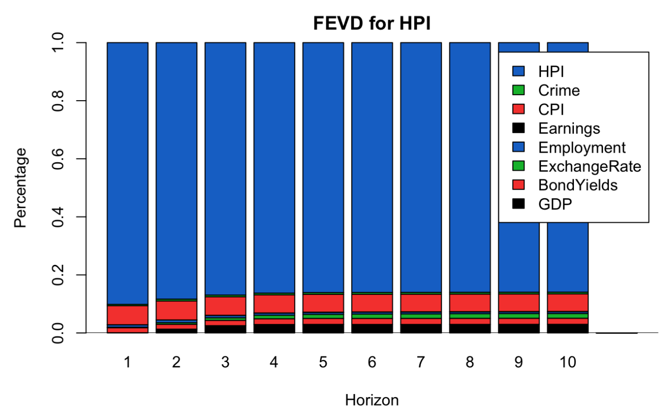

Figure 9 The forecast error variance decompositions (FEVD) for benchmark house prices.

The main takeaway from the decomposition is that house prices in Vancouver are largely explained by their own trend. That backs up the hypothesis from the Keynesian economists we mention in the introduction that people have strong expectations about future prices with little risk perceived. Therefore, the stronger the upwarding trend, the stronger their confidence in the market. Their confidence in the market leads them to buy more and the increase in demand pushes prices higher, which further increases their confidence.

Policy To assess the influence of Bill 28, We adjust initial VAR model by including the policy dummy variable as an exogenous variable.

In the new VAR model, the coefficient for policy in the Ordinary Linear Regression model of house price is -1.08, statistically significant at the 95% level. With the dependent variable measured as $100 \times \Delta \log(HPI_t)$, this implies the average monthly HPI growth decreases by *1.08 percentage points* during the policy period. In addition, it better reflects the role of foreign buyers as the relation between exchange rate and house price becomes significant at 95% significance level in Granger Causality Test and we observe that the role of domestic indicators such as GDP become weaker in explaining house price growth rate trend in the variance decomposition.

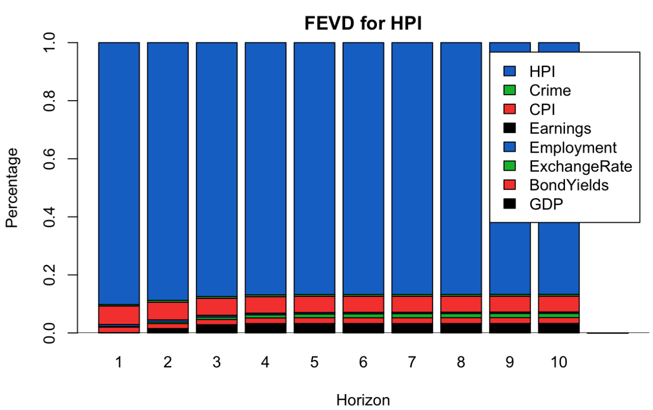

Figure 10 The forecast error variance decompositions (FEVD) for benchmark house prices (VAR model including policy).

Forecasting The following figure demonstrates the forecast result based on the VAR model. Based on the latest statistics for the first three months in 2022, it seems that VAR predicts better than ARIMA does as it succeeds in predicting the decrease in growth rate in the second month.

The following figure demonstrates the forecast result based on the VAR model with policy added, through which we can compare VAR forecasts with and without a policy to assess the effectiveness of the policy.

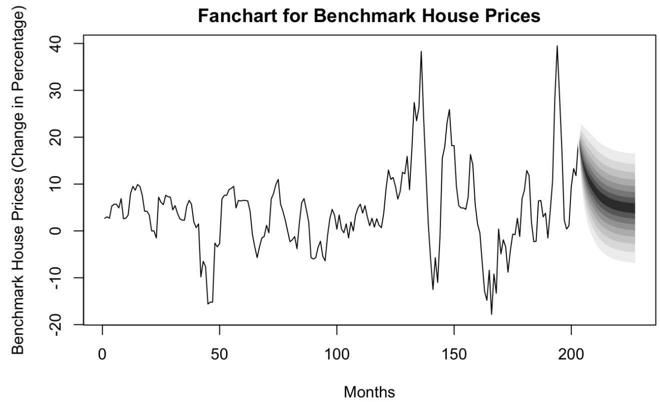

Figure 11 Fanchart for two years benchmark house prices forecasts based on a VAR model (without the policy).

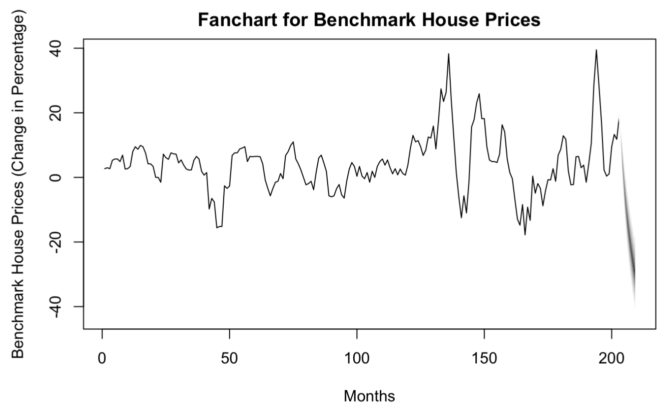

Figure 12 Fanchart for six months benchmark house prices forecasts based on a VAR model (with the policy).

The VAR with policy—with the effect of the tax on foreign purchases—predicts that house price growth will slow or even begin to decline after a few months, but the VAR without policy predicts that house price growth will still rise after 24 months. The effect of policy may be amplified by the model because we assume it is exogenous. However, it is hardly entirely exogenous.

In addition, as we elaborated above, sentiment/expectation plays an important role in the market. Our prediction is based on the previous observations. When the policy first comes out, people may pay more attention to it and overestimate its effect in the market. However, the policy came out in August 2016 and housing prices generally maintained a strong upward trend during the next three years. Though, statistically speaking, the policy decreases the growth rate of price when other variables are accounted for, regular buyers may not see the effect. They just see that prices keep going up.Therefore, it seems biased to use estimates about results of the policy to forecast housing price trends over a longer period.

4. Strength and Weakness

Strength

In this project, we conduct the time series analysis to present dynamic behavior between economic and social indicators and house price in Vancouver. We confirmed that the strength of macro indicators is weak without being trapped by endogeneity. We directly addressed the effect of the policy. We used statistical tools to shed light on social events and economic development and rationalize the statistics in a combination of economic theories and social studies to characterize the attributes of the housing market in Vancouver.

Hopefully, this study provides the audience with a deeper understanding of the development and future of the real estate market in Vancouver.

Weakness

There are certain weaknesses that come from our model. First, standard methods of statistical inference could yield biased results if some variables are highly persistent (Stock and Watson 2001). Without modification, standard VARs miss nonlinearities. In addition, in the impulse response analysis it is assumed that all effects of omitted variables are in the white noise (Kotzé 2022). A further robustness check is needed to ensure no highly influential variables are excluded from the system to ensure the validity for structural interpretation in impulse response. We did not sophisticatedly evaluate the precision of our forecasting results but mainly used them to analyze the social and economic implications. Furthermore, we did not expand on the robustness check part in the paper and examine our result in a larger context.

There are certain limitations that come from data sources; we were not able to find the monthly GDP, employment, weekly earnings data for the Greater Vancouver Area, and the monthly GDP for Canada. The monthly employment and weekly earning data for British Columbia were the only accessible sources for us, which may not truly reflect the economic situation in Greater Vancouver. For the crime data, we could only find the number of crimes recorded by the Vancouver police department that had occurred in the city of Vancouver, not in the whole Greater Vancouver area, which could also have impacts on our model's accuracy. In addition, the BC government expanded the rate of transaction tax on foreign buyers from 15% to 20% in February 2018. We assumed that the market was not sensitive to the 5% range and did not differentiate the effect of the change in our analysis.

5. Discussions and Conclusion

Our Story

Housing prices in Vancouver are expected to continue surging in the coming months after December 2021 based on both ARIMA and VAR, which is partly justified by the new housing price data we have collected in the first months of 2022. The effect of common macroeconomic indicators on Vancouver house price growth is limited, which is consistent with previous research such as Gordon (2020b). Although employment figures and other economic indicators, such as the GDP, can also independently provide statistically significant information about the future value of house prices based on Granger causality tests, their relative correlations remain weak. Housing prices respond to the impulse of GDP at period 3, which suggests a delayed effect. The market responds to bond yield (correlated to interest rate) and CPI instantly at period 1, at a significant level of 90%. The trend of housing prices is only most significantly explained by its own trends from different temporal junctures. Putting together with the variance decomposition plot, it is likely that purchasers buy on the assumption that house values will continue to rise and that the risk of loss is minimal, as they see its consistency predicts its continuity.

To significantly control housing prices and not risk bursting the bubble, a vigorous tax policy seems to be the most effective tool (Mhadi and Pinto (2018); Hu (2018)). Nevertheless, effects of the policy in the long run may remain to be further examined. Given the weak explanatory power of local economic indicators for the upward trend and the effectiveness of policies targeting foreign buyers, we can indirectly infer those foreign buyers may have played an important role in creating the housing bubble, which is also backed up by previous literature (Grigoryeva and Ley (2019); Moos and Skaburskis (2010)).

Placing our statistical results in the context of previous research and theories, we believe that market sentiments and strong expectations play crucial roles in maintaining housing bubbles.

6. Acknowledgments

We would like to express our profound gratitude to our Professor Heggeseth for her time and efforts throughout the semester. Your useful advice and suggestions were really helpful to us during the project's completion. We would also like to express our special thanks to Professor Kruger in the Department of Economics at Macalester College for his contribution to our knowledge of economics.

7. Conflicts of Interest

The authors declare no conflict of interest.

References

authority of the Minister. 2021. Guide to the survey of employment, payrolls and hours.

Becker GS. 1968. Crime and punishment: An economic approach. The Economic Dimensions of Crime. p. 13–68.

Case KE, Shiller RJ. 2003. Is there a bubble in the housing market? Brookings papers on economic activity. 2003:299–362.

CREA. 2016. Canadian real estate association.

Dreger C, Wolters J. 2008. Money velocity and asset prices in the euro area. SSRN Electronic Journal. .

Eviews. 2020. User's guide : Multiple equation analysis : Vector autoregression and error correction models.

FederalReserveBoard. 2022. Board of governors of the federal reserve system.

Fernando J. 2022. Consumer price index (cpi).

Freeman R. 1996. Why do so many young american men commit crimes and what might we do about it?. .

Gordon J. 2016. .

Gordon JC. 2020a. Reconnecting the housing market to the labour market: foreign ownership and housing affordability in urban canada. Canadian Public Policy. 46:1–22.

Gordon JC. 2020b. Solving puzzles in the canadian housing market: foreign ownership and de-coupling in toronto and vancouver. Housing Studies. pp. 1–24.

Government of Canada SC. 2022a. Gross domestic product by industry - national (monthly) (gdp).

Government of Canada SC. 2022b. Survey of employment, payrolls and hours (seph).

Grigoryeva I, Ley D. 2019. The price ripple effect in the vancouver housing market. Urban Geography. 40:1168–1190.

Hu J. 2018. The effects of british columbia's vacancy tax and foreign-buyer tax act on the supply of new residential housing in vancouver. .

Keynes JM. 1936. *The General Theory of Employment, Interest and Money*. Macmillan. London.

Kotzé K. 2022. .

Lazzeri F. 2021. *Machine learning for time series forecasting with python*. John Wiley Sons, Inc.

Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis*. Springer. Berlin.

Mhadi H, Pinto J. 2018. Assessing the effects of the vancouver foreign-buyer transfer tax on housing prices and volume. Western Undergraduate Economics Review. 17:1–22.

Moos M, Skaburskis A. 2010. The globalization of urban housing markets: Immigration and changing housing demand in vancouver. Urban Geography. 31:724–749.

Paparoditis E, Politis DN. 2016. The asymptotic size and power of the augmented dickey–fuller test for a unit root. Econometric Reviews. 37:955–973.

Penn State. 2022. Vector autoregressive models var(p) models.

REALTORS. 2022. Market stats.

Salvi J. 2019. Significance of acf and pacf plots in time series analysis.

Sims C. 1986. Are forecasting models usable for policy analysis? Quarterly Review. 10.

Solow RM. 1956. A contribution to the theory of economic growth. The Quarterly Journal of Economics. 70:65.

Stock JH, Watson MW. 2001. Vector autoregressions. Journal of Economic Perspectives. 15:101–115.

Todd D. 2017. Vancouver is the most 'asian'city outside asia. Vancouver Sun [Internet]. .

Wikipedia contributors. 2021. Bill 28 (british columbia) — Wikipedia, the free encyclopedia. [Online; accessed 9-May-2022].

Zhu X. 2022. Forecasting and time series methods notes.

Zivot E, Wang J. 2006. Vector autoregressive models for multivariate time series.