STA302 Fall 2024 Final Project Part 1: Research Proposal and Data Introduction

Research Proposal and Data Introduction

Wendy Huang Gadiel David Flores

2024-07-10

1 Introduction

Housing affordability and price fluctuations have become pressing issues in metropolitan areas, and Toronto is no exception. As one of Canada's largest and fastest-growing cities, understanding the factors influencing house prices in Toronto is crucial for policymakers, developers, and potential homeowners. Over the past decade, housing prices have seen a steep upward trend, driven by a combination of housing supply and demand, economic, and policy-related factors.(Scrinko (2024)) This study aims to investigate the relationship between several key housing market indicators—namely, the number of available homes, homes sold, homes starting construction, new homes completed, and interest rate hikes—and housing prices in Toronto. By analyzing these factors, this research will contribute to a more comprehensive understanding of Toronto's housing market dynamics.

The rapid rise in housing prices has been widely debated, with supply constraints being one of the leading factors. For example, Glaeser, E. L., Gyourko, J., & Saks, R. E. (Edward L. Glaeser (2005)) employed linear regression to analyze the relationship between supply constraints and prices, showing an imbalance between supply and demand results in price hikes. This paper as well supports the use of housing starts and completions as critical indicators of price trends, aligning with this proposed research focus on construction-related factors in Toronto's housing market.

Additional studies support the notion that housing availability and construction trends influence prices. In their paper, Goodman, A. C., & Thibodeau, T. G. (Allen C. Goodmana (2007)) explored the relationship between new housing construction and price stability in urban areas. The authors found that increased housing completions have a stabilizing effect on prices, especially in high-demand areas. Their methodology relied on linear regression, confirming its suitability for modeling how newly completed homes impact price trends.

The paper by Guiwen Bai and Yu Ma (Yu Ma (2017)) uses a multi-factor regression model to study how these fundamentals, particularly the cost of borrowing and income tax impacts and interest rates, relate to housing price movements in Canada. The paper confirmed that low interest rates contribute to increased affordability, encouraging more people to enter the housing market.

Given these findings, linear regression is an appropriate tool for analyzing how these supply-side continuous variables interact to affect prices in Toronto. The clear quantitative relationship between supply predictors (available homes, homes sold, etc.) and price response aligns dynamically well with the assumptions of linear regression, making it a suitable method for this analysis. Linear regression allows us to quantify the relationship between these factors and price, providing insights into which variables have the most significant impact.

2 Data Description

2.1 Data Source

Using data from Statistics Canada, we extracted five datasets. Statistics Canada (2024e) served as the foundation for this research paper, as it contained the main statistic we wanted to analyze. Statistics Canada (2024e) represents the new housing price index, which calculates the average cost of new housing each month by multiplying by 100 and dividing by the base year. This provides an index reflecting housing costs for any given year and month. We then utilized Statistics Canada (2024c), Statistics Canada (2024d), Statistics Canada (2024b), Statistics Canada (2024g), and Statistics Canada (2024a) to extract our 13 predictors, including Absorption, GDP, CPI, interest rates, and Construction. Specifically, Absorption and Construction are further divided into two parts. Absorbtion is divided into semi-detached homes and single homes as well as empty and sold houses for any given month. Construction is categorized into three parts: new construction, under construction, and finished construction. These datasets were cleaned by removing missing values, converting data into whole integers, and merging them into one dataset. Additionally, the data spans from 1997 to 2016 and is divided quarterly, providing 76 entries. Lastly, we used interest rates to create our categorical predictor. This was done by comparing the entry i with entry i+1 in order to determine if there was an interest hike.

2.2 Response and Predictor Variables

Given our research question how do supply-side factors, such as housing availability, sales volume, construction starts, completions, and interest rate, impact housing prices in Toronto, the response variable chosen to help answer our research question is 'new housing price index'. This variable is suitable for linear regression because it is continuous, enabling us to model its relationships with multiple predictors. Additionally, since the new housing price index is

influenced by economic variables, it allows us to identify trends and patterns through our linear regression model. For these reasons, it is a strong candidate for our model. Our five predictor variables are the number of available homes, homes sold, homes starting construction, new homes completed, and rate hikes

3 Ethics Discussion

The datasets extracted from Statistics Canada are both trustworthy and ethically collected, as they originate from a reputable national statistical agency committed to maintaining data integrity and transparency. Statistics Canada follows strict protocols for data collection, ensuring that the information is accurate, reliable, and reflective of real market conditions (Statistics Canada (2024f)). While the datasets themselves are typically reliable and collected following rigorous ethical standards, potential risks related to racial and ethnic disparities in housing must be acknowledged.

Ethical considerations also arise from the treatment of sensitive information, particularly regarding interest rates and housing data, which could impact individual financial circumstances. The datasets used in this study focus on aggregate data rather than individual responses, thus minimizing privacy concerns.

Moreover, the datasets underwent thorough cleaning processes to ensure accuracy, further reinforcing their reliability for analysis. Overall, the ethical collection practices and transparency associated with these datasets support their trustworthiness for examining the impacts of supply-side factors on housing prices in Toronto.

4 Preliminary Results

Our model demonstrates the relationship between the New Housing Price Index (NHPI) and various predictors, such as detached absorption and unabsorbed homes, homes starting construction, completed constructions, and rate hikes. To start, we will discuss the intercept, which is 88.52 when all predictors are zero. This serves as a baseline for evaluating the effects of other predictors.

The first predictor is detached absorption. For each additional home sold in a quarter, the expected decrease in the NHPI is 0.0227 points. In contrast, unabsorbed homes cause an increase of 0.005315 points in the NHPI. New homes beginning construction have a minimal impact, with a 0.0009611 point increase in the NHPI for every home that starts construction. Meanwhile, completed homes have a greater impact, with a 0.002977 point increase. Lastly, rate hikes have the most significant impact: when there is a rate increase, the NHPI decreases by 13.41 points.

4.1 Residual Analysis

In this section, we will explore the assumptions made for our linear regression models and check for any violations. To begin, using our model in Section 4.2, we created residual plots to verify these assumptions. The first graph we will examine is Figure 2, which is our Residuals vs. Fitted Values plot. This graph shows some clustering around a fitted value of 100 and above the 0 residual line. It also displays a decreasing fanning pattern. These two characteristics suggest potential non-linearity and some correlated errors.

Next, we will review the residuals vs. predictors. In Figure 5 and Figure 6 The plots for homes being built and completed homes show the least amount of assumption violations, as the graphs exhibit good spread and randomness. In contrast, the plots for homes sold Figure 3 and unsold homes Figure 4 display some fanning and clustering, which may indicate correlated errors and non-constant variance. Our categorical predictor, rate hike Figure 7 , reveals some non-constant variance, as the spread of residuals for "1" is wider than for "0."

Lastly, while most residuals align well with the line in the QQ plot Figure 8 , there are deviations at the lower and upper extremes, suggesting some skewness. This indicates that the residuals are not perfectly normally distributed, which may signal a violation of the normality assumption.

4.2 Model Fitting

Data
- **Quarterly Average**: This is our response variable, representing the average cost of a home in Canada.
- **Detached Absorption Quarterly Avg**: This is our first predictor variable, describing the number of houses sold in a quarter.
- **Starting Detached Construction**: This predictor variable indicates the number of houses that began construction in a quarter.
- **Completed Construction Detached**: This represents the number of homes completed in a given quarter.
- **Rates**: This is our categorical predictor, with a value of 1 if there was a rate hike in that quarter or 0 if there was not.
- **Detached Unabsorbed Quarterly Avg**: This is our final predictor, indicating the number of available houses that were not sold during that quarter.

Table 1: Summary Table of Regression Coefficients

Variable	Coefficient
(Intercept)	88.5220572
Detached_Absorption_Quarterly_Avg	-0.0227041
Detached_Unabsorbed_Quarterly_Avg	0.0053151
Starting_Detached_Construction	0.0009611
Completed_Construction_Detached	0.0029772
rates	-13.4055414

Figure 1: Multiple Linear Model

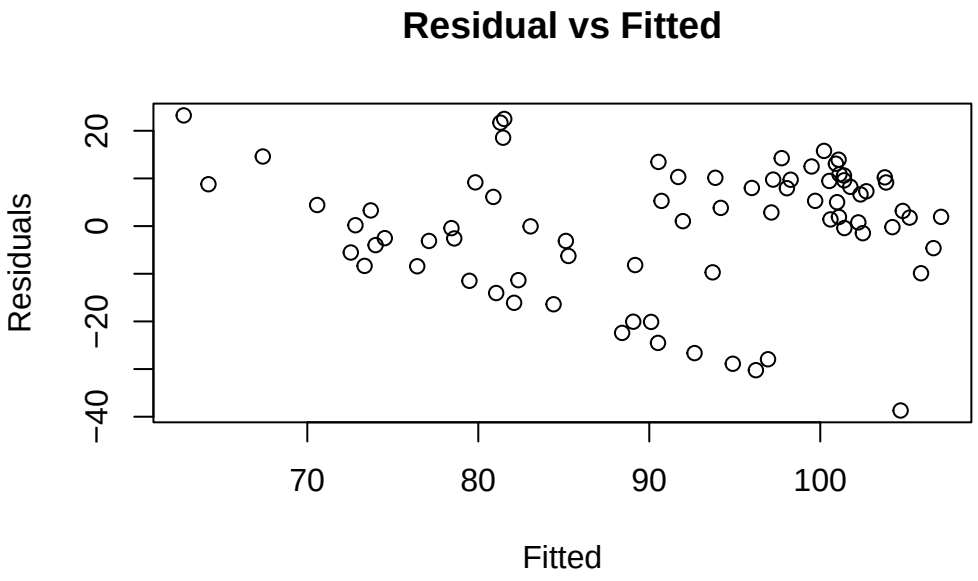

Figure 2: Residual vs Fitted Values

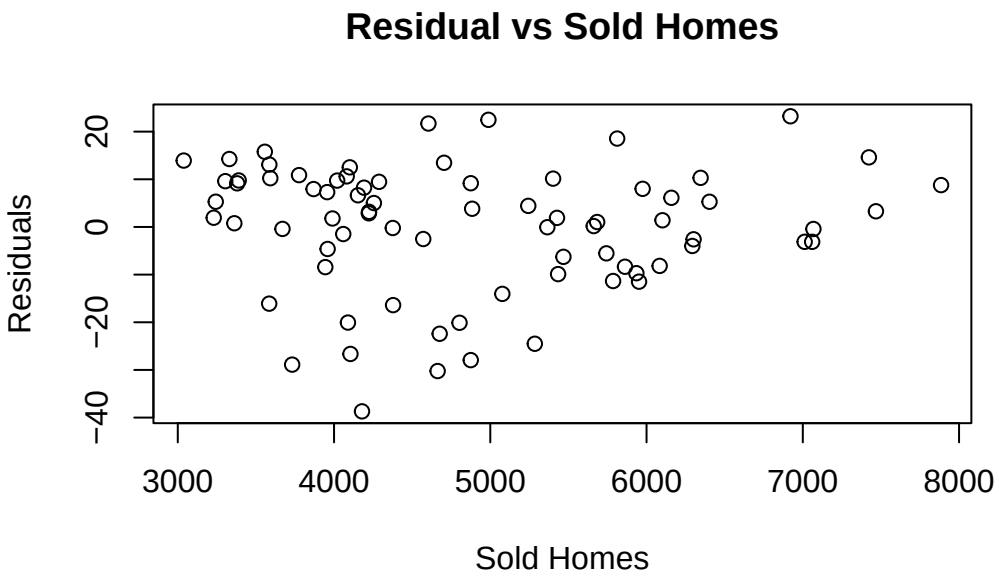

Figure 3: Residuals vs Each Predictor

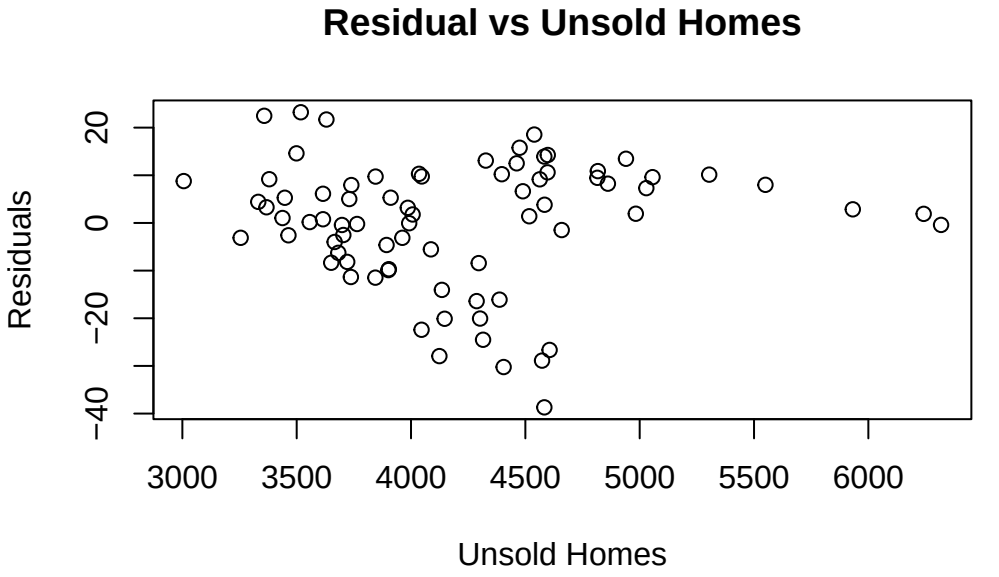

Figure 4: Residuals vs Each Predictor

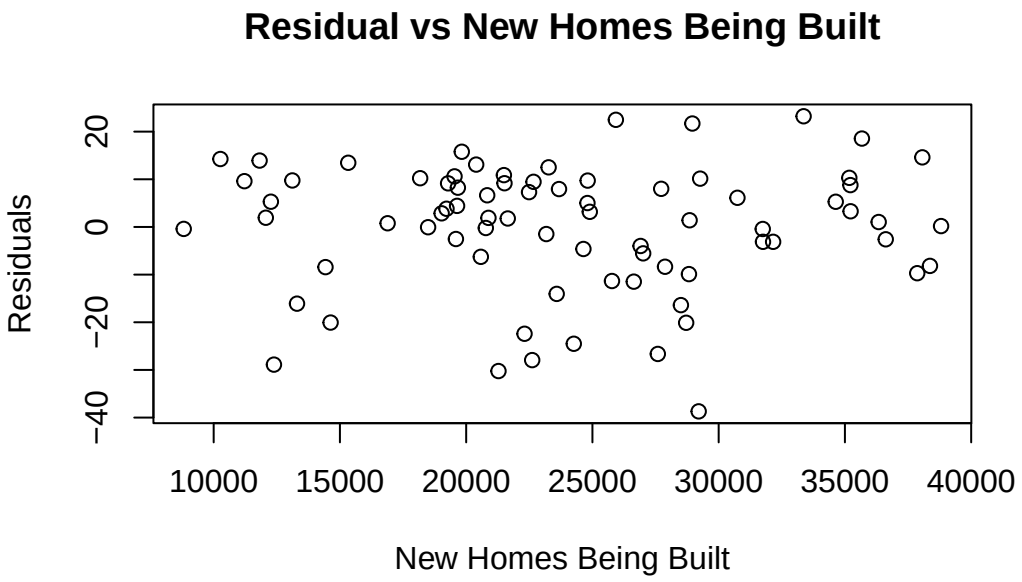

Figure 5: Residuals vs Each Predictor

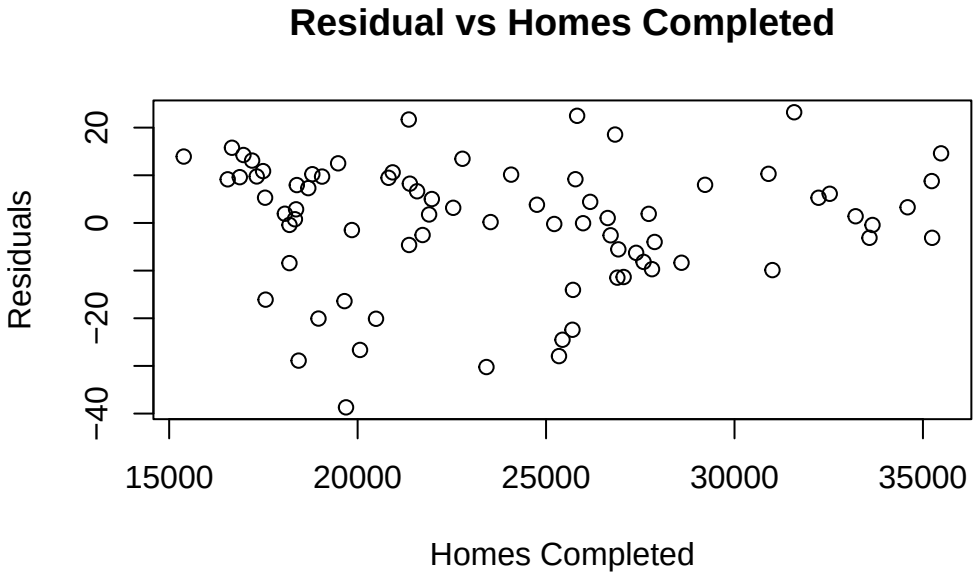

Figure 6: Residuals vs Each Predictor

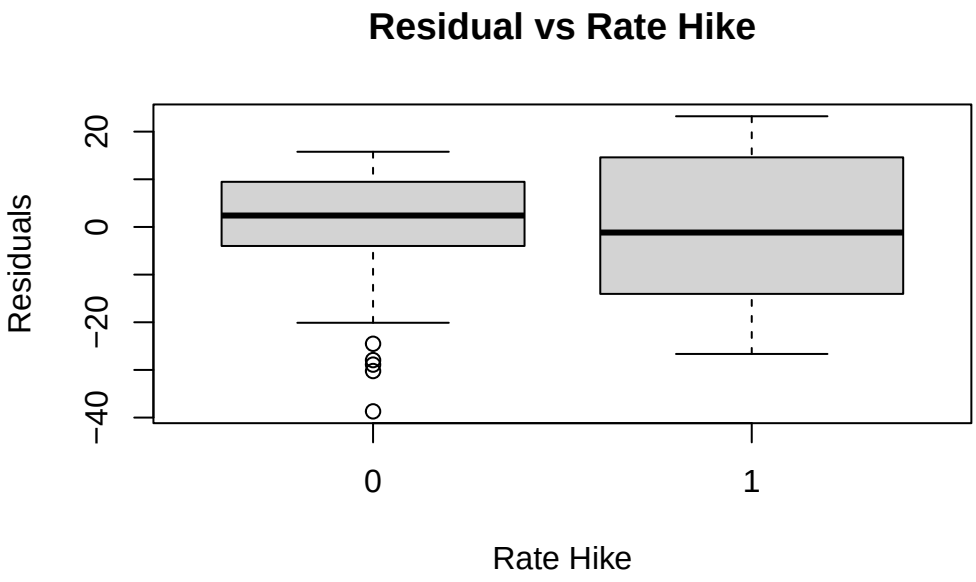

Figure 7: Residuals vs Each Predictor

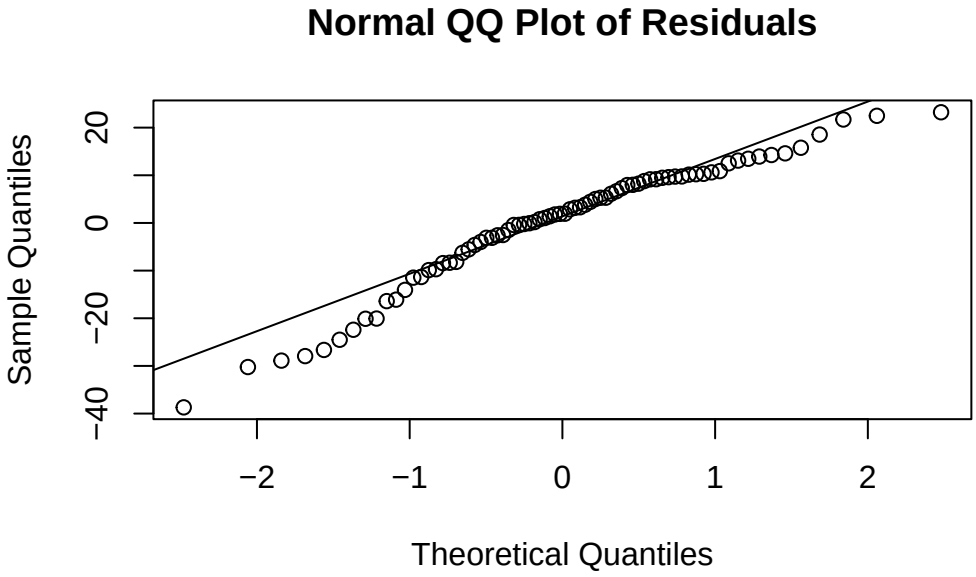

Figure 8: QQ Plot

Bibliography

Allen C. Goodmana, Thomas G. Thibodeau. 2007. "The Spatial Proximity of Metropolitan Area Housing Submarkets." *Real Estimate Economics*. https://doi.org/https://doi.org/10.1111/j.1540-6229.2007.00188.x.

Edward L. Glaeser, Joseph Gyourko & Raven E. Saks. 2005. "Why Have Housing Prices Gone Up?" *National Bureau of Economic Research*. https://doi.org/10.3386/w11129.

Scrinko, Jordon. 2024. *5 Causes Behind the Toronto Housing Crisis*. https://precondo.ca/toronto-housing-crisis/.

Statistics Canada. 2024a. *Canada Mortgage and Housing Corporation, Absorptions and Unabsorbed Inventory, Newly Completed Dwellings, by Type of Dwelling Unit in Census Metropolitan Areas*. Statistics Canada. https://doi.org/https://doi.org/10.25318/3410014901-eng.

———. 2024b. *Canada Mortgage and Housing Corporation, Housing Starts, Under Construction and Completions, All Areas, Quarterly*. Statistics Canada. https://doi.org/https://doi.org/10.25318/3410013501-eng.

———. 2024c. *Consumer Price Index, Monthly, Not Seasonally Adjusted*. Statistics Canada. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000401.

———. 2024d. *Gross Domestic Product (GDP) at Basic Prices, by Industry, Monthly (x 1,000,000)*. Statistics Canada. https://doi.org/https://doi.org/10.25318/3610043401-eng.

———. 2024e. *New Housing Price Index (2007=100)*. Statistics Canada. https://doi.org/https://doi.org/10.25318/1810005201-eng.

———. 2024f. *Statistics Canada*. https://www.statcan.gc.ca/en/start.

———. 2024g. *Table 10-10-0122-01 Financial Market Statistics, Last Wednesday Unless Otherwise Stated, Bank of Canada*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1010012201.

Yu Ma, Guiwen Bai. 2017. "FUNDAMENTALS AFFECTING CANADIAN HOUSING MARKET." chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://summit.sfu.ca/_flysystem/fedora/2022-08/input_data/nid_17821/Bai%252C%2520G%2520%252526%2520Ma%252C%2520Y.pdf.