My title*

My subtitle if needed

Krishna Kumar

November 24, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

This report uses monthly economic data derived from the Federal Reserve Economic Data (FRED) database to examine the key factors influencing housing affordability in the United States of America from March 2006 to September 2024. For this analysis, I have combined five data sets into one, which is then used for analysis. The data sets were downloaded using the `fredr` **(CITE)** package and loaded using the same library. The data were cleaned and analysed using the statistical programming language `R`. Simulating, cleaning, testing, and graphing were done with the help of the following packages: `Tidyverse`, `dplyr`, `lubridate`, `testthat`, `readr`, `arrow`, `ggplot2`, `here`, **(CITE + ADD ANY MORE YOU USED HERE)** this

---

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

1

The Federal Research Economic Data provides information about the people and economy of the United States, with a goal to support economic growth, assist in making informed decisions, and enhance scientific knowledge. Housing affordability has emerged as a growing concern in the US, with people facing a shortage of houses and constantly increasing prices due to a variety of factors. These challenges, compounded by the COVID-19 pandemic, have left many families struggling to access affordable housing. **(CITE 1) FROM REFERENCES)** Studying this data is essential in assessing the economic factors affecting housing affordability as well as the financial wellbeing of individuals. The data sets were all downloaded separately using the FRED API and then combined during the data cleaning.

2.2 Measurement (PARTIALLY LEFT TO DO)

For each variable:

- ***CPI***: The CPI is based on prices for food, clothing, shelter, and fuels; transportation fares; service fees (e.g., water and sewer service); and sales taxes. Prices are collected monthly from about 4,000 housing units and approximately 26,000 retail establishments across 87 urban areas. To calculate the index, price changes are averaged with weights representing their importance in the spending of the particular group. The index measures price changes (as a percent change) from a predetermined reference date.

- ***Interest Rate:*** The Federal Open Market Committee (FOMC) meets eight times a year to determine the federal funds target rate. These rates are then announced to the public.

- ***Average Wage:*** **Have to go through website for this**

- `House Price`: **No method of measurement given for this.**

- `House Supply`: **No method of measurement given for this.**

2.3 Data Cleaning

The main aspect of the data cleaning was merging all five data sets into one data set. Moreover, the columns were renamed to remove any informal names and two new columns were created using the existing data sets: House Prices Growth (%) and Wage Growth (%). This was done using the simple formula of (House Price 2 - House Price 1)/House Price 1, calculating the percentage change. The same was done for average wages. This allowed us to analyse the growth rates year over year and get a sense of how fast each variable was growing through time. Additionally, another column was created titled 'Affordability Index.' This column was created by dividing the values of House Prices by the Average Wage in the same rows. This gave us an outcome of the number of hours an individual would have to work to buy a house outright, serving as a good indicator of housing affordability, especially when looking at the data through a long-term lens.

2.4 Data Limitations and Similar Data sets

While the data from FRED provides valuable insights into macroeconomic variables, it is not without limitations. Notably, the data on wages reflects average hourly earnings of all employees in the private sector, rather than total household income, which may overlook other forms of income, such as government transfers, investments, or multi-income households, that influence affordability. Additionally, the data is primarily aggregated at the national level, limiting its ability to capture trends at the regional level in housing affordability, which can be substantial across states and metropolitan areas. Another consideration is that some variables are seasonally adjusted while others are not. However, over a long period of time, like the nearly two-decades used in this analysis, these seasonal adjustments are unlikely to impact the broader trends and implications this paper aims to study.

Similar data sets for all variables were readily available on the FRED database, along with other databases like the U.S. Census Bureau or Bureau of Labor Statistics. While those data sets might be able to provide regional or demographic data, the chosen variables were selected due to the monthly recording of data for a more detailed analysis. Despite these limitations, the FRED data sets turned out to be the best choice for an analysis fo this kind.

2.5 Variables of Interest

For analysis, only the required variables' data sets were downloaded from FRED. This paper decided to focus on five variables to assess the problem of unaffordable housing in the US: CPI, interest rate, housing supply, median housing prices, and average wages. They are described below:

- ***CPI***: This variable represents the Consumer Price Index for All Urban Consumers, for all items in the U.S. City Average, which measures the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services. Recorded monthly, this variable is measured in an index form, for which the base period is 1982-1984 where the value of the index sat at 100. Change in CPI from last year's CPI is one of the most common measures used to calculate inflation. This variable is seasonally adjusted and looks at ~ 88% of the population. **(CITE 2) HERE FROM REFERENCES)**

- ***Interest Rate:*** This variable refers to the Federal Funds Effective Rate, which is the interest rate at which depository institutions trade federal funds with each other overnight. This interest rate is the central interest rate in the US, influencing other interest rates such as the prime rate (what banks charge customers with higher credit ratings) and longer term rates (mortgages, loans, savings). This variable is recorded monthly, represented as a percentage, and is not seasonally adjusted. **(CITE 3) HERE FROM REFERENCES)**

- *Average Wage:* This variable denotes the Average Hourly Earnings of All Employees in the Total Private Sector. It includes premium pay for overtime and late-shift work. This was chosen over wage rates as this measures the actual return to a worker for a set period of time. It excludes benefits, irregular bonuses, retroactive pay, and payroll taxes paid by employers. This variable is measured in US Dollars per hour, recorded monthly, and is seasonally adjusted. **(CITE 4) FROM REFERENCES)**

- *House Price:* This variable captures the Media Sales Price for New houses Sold int he United States. The media was chosen over the average prices to remove the effects of any outliers. This variable is recorded monthly, measured in US dollars, and is not seasonally adjusted. **(CITE 5) FROM REFERENCES)**

- *House Supply:* This variables measures the Monthly Supply of New Houses in the United States. Measured in 'month's supply,' this refers to the ratio of new houses for sale to new houses sold. The month's supply indicates the size of the new for-sale inventory compared to the number of new houses currently being sold. This variable is recorded monthly and is seasonally adjusted. **(CITE 6) FROM REFERENCES)**

- *House Price Growth:* This variable represents the year-over-year percentage growth in the Median Sales Price for New Houses Sold in the United States. It is calculated by taking the difference between the median house price at the end of the year and the beginning of the year, dividing this difference by the value at the beginning of the year, and multiplying by 100 to express the result as a percentage. This highlights the annual rate of change in house prices, helping to track trends in housing market inflation or deflation.

- *Wage Rate Growth:* This variable reflects the year-over-year percentage growth in the Average Hourly Earnings of All Employees in the Total Private Sector. The calculation follows the same approach as house price growth: the difference between wages at the end of the year and the beginning of the year is divided by the wage value at the start of the year, then multiplied by 100 to express the result as a percentage. This variable captures how wages change annually, offering insights into trends in earning power.

- *Affordability Index:* This variable measures the ratio of the Median Sales Price for New Houses Sold in the United States to the Average Hourly Earnings of All Employees in the Total Private Sector. It represents how many hours of work are required to afford a median-priced home. A higher value indicates reduced affordability, as workers would need to work more hours to buy a home.

2.5.1 Consumer Price Index Trends Over Time

Figure **(CROSS REFERENCE HERE)** shows a clear pattern of a gradual increase from 2006 to 2024, with a clear spike in 2007 due to the financial crisis and a drop immediately after 2020 as a result of the COVID-19 pandemic. We can also see a notable acceleration in
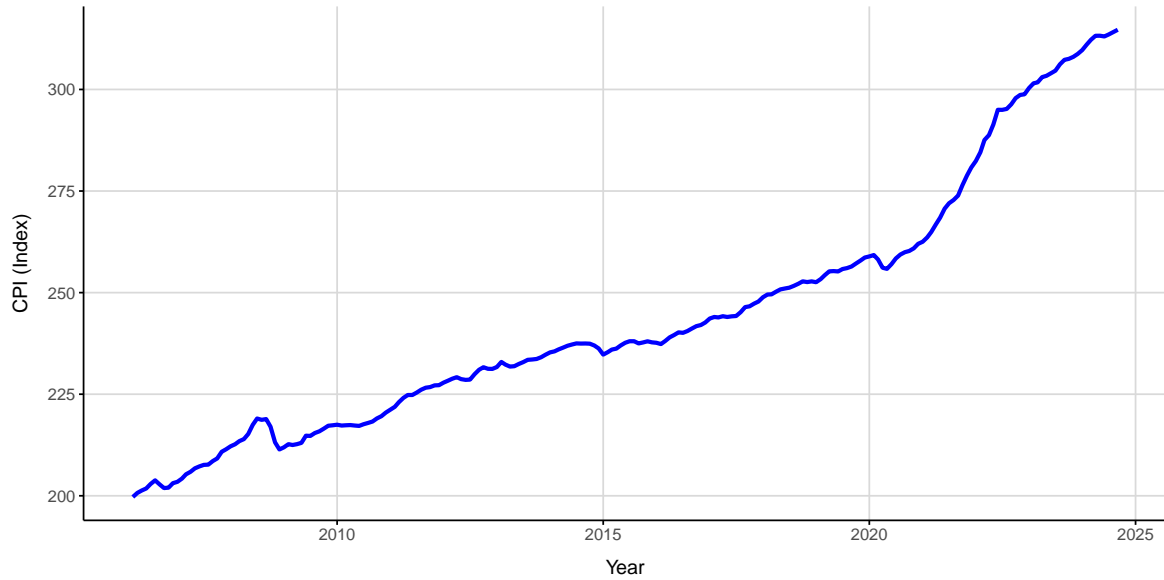
4

Figure 1: Consumer Price Index (CPI) Over Time

the post pandemic period CPI where the trend line becomes much steeper. This highlights periods of economic expansion, but also signals rising costs for consumers.

Inflation, as measured by CPI, directly impacts housing affordability. As inflation rises, the cost of living increases, including housing prices. Higher inflation often prompts the Federal Reserve to raise interest rates (which we will see in later graphs), making mortgages more expensive and reducing affordability for potential home buyers. Furthermore, when inflation erodes purchasing power, individuals have less disposable income to allocate toward homeownership, exacerbating affordability challenges. Historical relationships between CPI and housing prices reveal that rising inflation can contribute to increased home prices due to higher costs for construction materials and labor. **(CITE 7) HERE FROM REFERENCES)**

2.5.2 Interest Rates Trends Over Time
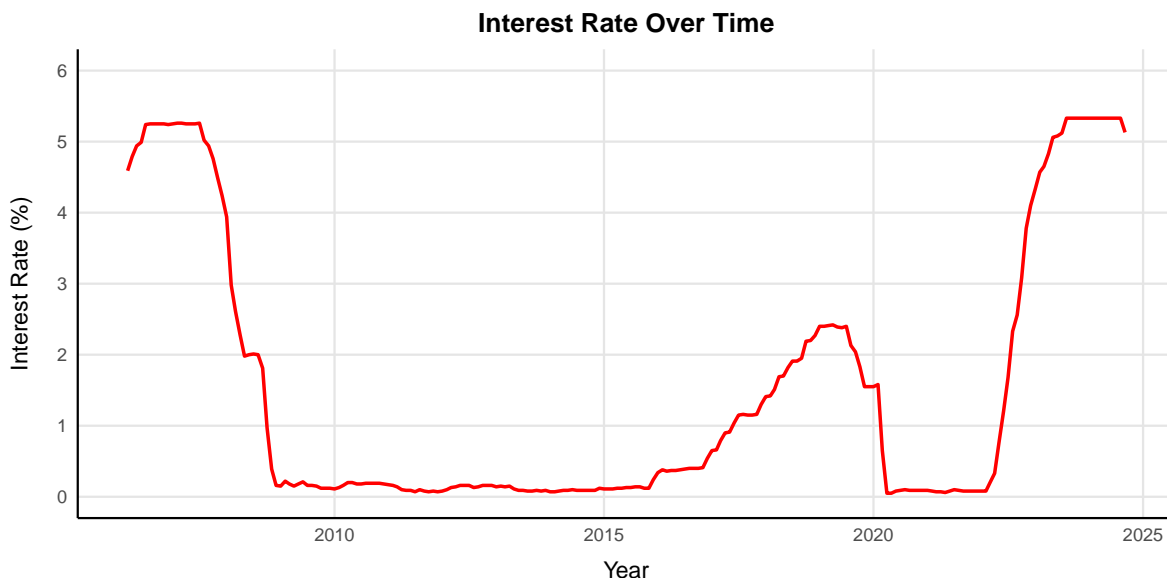
Interest Rate Over Time

Figure 2: Federal Funds Effective Interest Rate Over Time

Figure Figure 2 shows the irregular trends of the interest rate from 2006 to 2024. The interest rates saw a drastic drop during the financial crisis which began in 2007. From 2008 to 2015, interest rates remained historically low, hovering near zero due to monetary policy responses to the Great Recession. A gradual increase began in 2016, reflecting economic recovery and attempts to normalize monetary policy, before another sharp decline in 2019 as the COVID-19 pandemic spurred aggressive monetary easing to support the economy. Post-2022, we can see a dramatic rise in interest rates, corresponding to central banks' efforts to combat inflationary pressures. This sharp increase represents one of the fastest monetary tightening cycles in recent history.

Regarding housing affordability, these trends are crucial. Low interest rates, such as those seen during 2008–2015 and 2020–2021, generally reduce borrowing costs, making mortgages more affordable and stimulating demand in the housing market. Ironically, as we will see in Figure 3, during these times when borrowing becomes cheaper, we have seen housing supply fall drastically. Moreover, the sharp increases in interest rates post-2022 significantly raise mortgage rates, reducing affordability for new buyers and potentially cooling housing market activity, and of course, in these times, we have seen an increase in housing supply. These trends underline the strong interplay between monetary policy and housing market dynamics, highlighting the importance of interest rate policies in influencing housing affordability over time
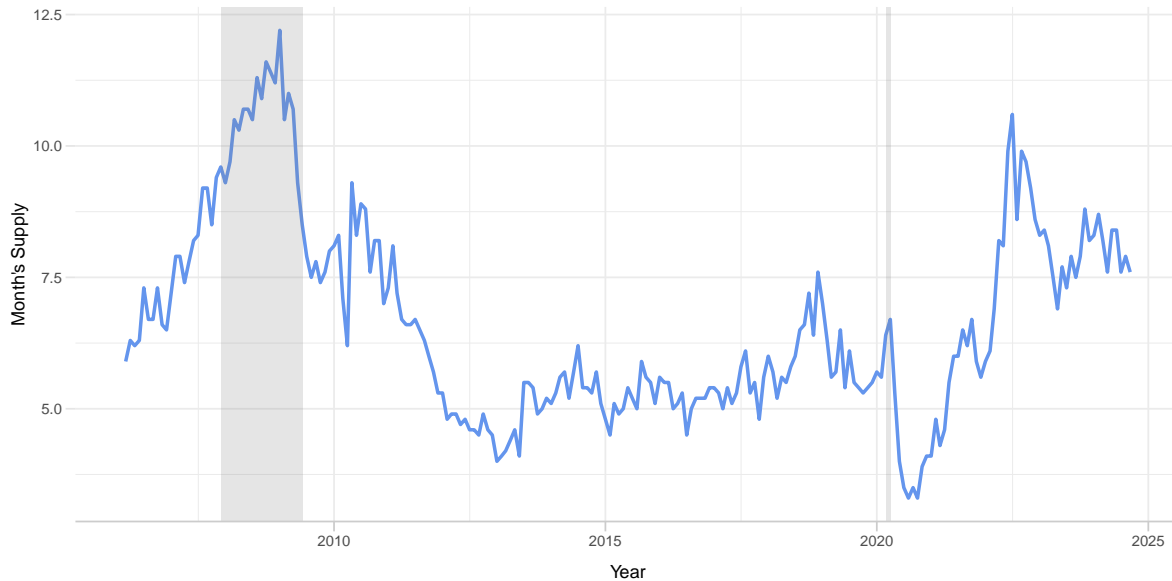
2.5.3 Housing Supply Over Time

Figure 3: Housing Supply Over Time

Figure Figure 3 highlights sharp fluctuations in housing supply in the past 19 years, with recessions marked by the shaded areas. In the aftermath of the 2007–2008 financial crisis, housing supply spiked to over 12 months, reflecting a surplus of available homes due to decreased demand and widespread foreclosures. However, as the economy began recovering, supply steadily declined, hitting a low point around 2013. This downward trend in supply coincided with robust demand, driven by low interest rates and favorable borrowing conditions.

During the COVID-19 pandemic, a combination of supply chain disruptions, labor shortages, and heightened demand for housing led to a sharp decline in supply, reaching historic lows. Post-2020, the supply began to rise again, likely due to reduced affordability as a result of rising interest rates (as seen in Figure Figure 2) and cooling housing market activity. However, it reached a high point in the first half of 2022 and then kept decreasing drastically till 2024.
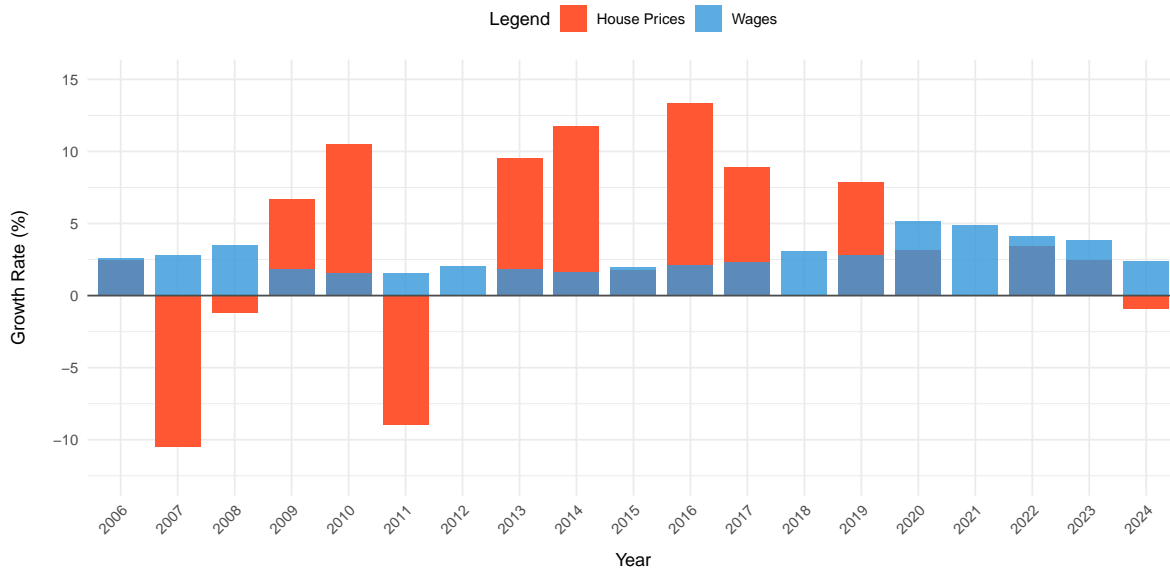
2.5.4 Wage Growth vs House Prices

Figure 4: Growth Rates of House Prices and Wages

Figure Figure 4 compares the annual growth rates of house prices and wages from 2006 to 2024. The data highlights a persistent gap between the two metrics, with house prices outpacing wage growth in 5 of the 19 years analysed. During the financial crisis of 2007–2008, house prices saw a sharp drop, with growth rates falling below -10%, while wage growth remained relatively stable. Post-2011, house prices recovered, showing strong growth from 2012 to 2016, significantly exceeding wage growth during this period. This divergence reflects a period of increasing housing demand, driven by low interest rates and insufficient housing supply, as seen in Figure Figure 3.

Notably, wage growth has consistently lagged behind house price growth, contributing to worsening housing affordability over the years. For example, during the COVID-19 pandemic in 2019, while wage growth remained positive, house prices surged, reflecting the impact of supply shortages and pandemic-induced demand. Post-2022, a notable shift occurred with house price growthdeclining due to rising interest rates and reduced affordability, while wage growth maintained a steady pace. These trends emphasize the widening affordability gap and the complex interplay between wages, housing prices, and macroeconomic conditions, highlighting the challenges faced by households in balancing income growth with escalating housing costs.
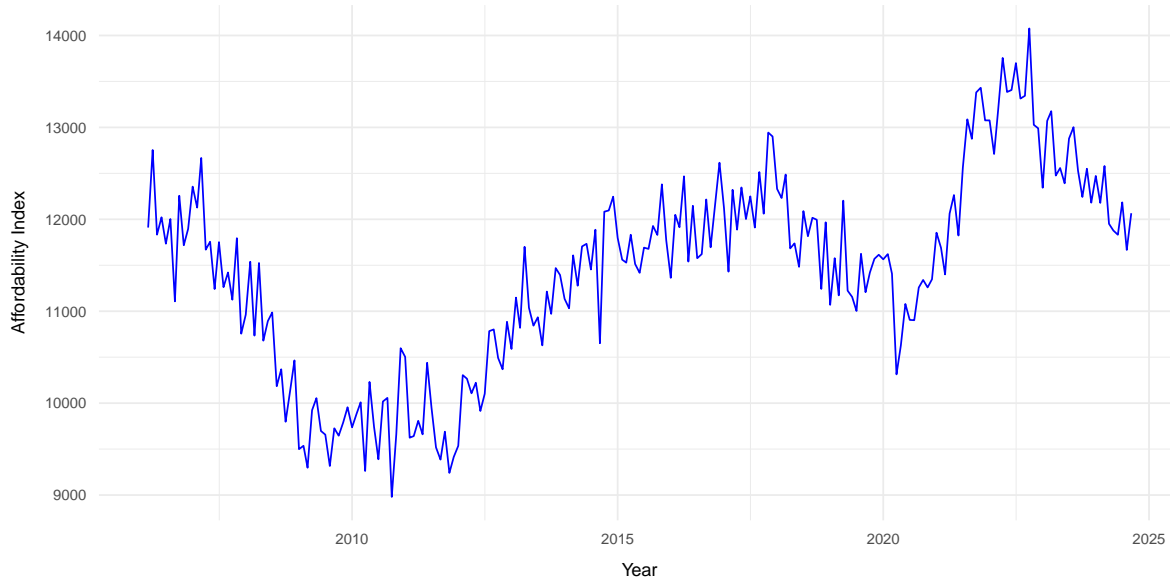
Figure 5: Growth Rates of House Prices and Wages

2.5.5 Affordability Index

Figure Figure 5 presents the trends in the affordability index from 2006 to 2024, illustrating fluctuations in the cost of housing relative to wages. The affordability index is a key metric that represents the number of work hours required to afford a house, with higher values indicating greater financial strain on individuals. The index shows notable volatility over time, particularly during key economic events. From 2008 to 2011, the index declined significantly, reflecting improved affordability during the global financial crisis, likely due to falling house prices and stabilizing wages. However, affordability worsened from 2012 onwards, as the index steadily increased, peaking around 2020. This peak coincides with rising house prices during the pandemic-driven housing boom, requiring more work hours to afford a home. Post-2020, the index shows a downward trend, reflecting corrections in housing prices amid rising interest rates.

3 Model

In this section, we develop and evaluate a series of models to understand the factors influencing housing affordability. The major challenge lies in selecting a model which is balanced in complexity and provides insightful results. We will be analyzing how CPI, housing supply, and interest rates affect the housing affordability in three separate models, adding one after the other. The goal of out modelling strategy is to find the effect of each of the predictor variables on the affordability index and their strengths.

Because housing prices or wage rates alone can't determine affordability, we created an affordability index to better measure affordability. This index has a simply formula where we divide housing prices by average wages, leading to a number that represents how many hours an individual would need to work to buy a house outright.

We begin with a single predictor, CPI, to assess it's impact on housing affordability. Then, we add housing supply for model 2 and interest rates for model 3, arriving at a full model that shows us the effect of each of these variables marginally, as well as collectively. This step wise approach helps ensure that the model remains interpretable. Moreover, performance measures such as R2, Adjusted R^2, AIC, BIC, and RMSE were used to evaluate model performance.

Here we briefly describe the multiple linear regression model used to investigate... Background details and diagnostics are included in Appendix B.

3.1 Model set-up

We aim to model the **Affordability Index**, which represents the number of hours of work required to afford a house, as a function of key economic variables such as interest rates and inflation (measured by CPI).

$$y_i = y_i = \beta_0 + \beta_1 \cdot \text{CPI}_i + \beta_2 \cdot \text{HouseSupply}_i + \beta_3 \cdot \text{InterestRate}_i + \epsilon_i$$

Where:

- y_i is the affordability index for observation iii, which measures the ratio of house prices to average wages, indicating housing affordability.

- β_0 is the intercept, representing the expected value of the affordability index when all predictors are zero.

- β_1 captures the effect of the CPI (Consumer Price Index) on the affordability index, reflecting inflation.

- β_2 captures the effect of the Housing Supply on the affordability index.

- β_3 captures the effect of the Interest Rate on the affordability index

- ϵ_i represents the error term, assumed to follow a normal distribution with mean 0, accounting for unexplained variability in the affordability index.

3.2 Model Justification

The linear regression model was chosen for this analysis to understand how each predictor variable affects affordability of housing. The affordability index is key to understanding patterns of affordability in the housing market; This is something average wages or house prices cannot determine along. CPI, housing supply, and interest rates all have impacts on housing affordability we expect to see a strong positive correlation between the affordability index and CPI and interest rates. Meaning that as these variables increase, the affordability index increases as well, indicating worsening affordability. However, we expect to see a negative relationship between the affordability index and housing supply. We also aim to measure which of these variables has a stronger impact on housing affordability.

An increase in interest rates increases the mortgage rates in turn and decreases housing affordability. An increase in CPI represents an increase in inflation, which then usually leads to an increase in housing prices. Lastly, a decrease in the housing supply leads to a shortage of homes and therefore, leads to increased housing prices and lower affordability. These factors work alone and together, to influence the financial burden on individuals. We decided not to choose wage growth or house price growth as predictor variables in our analysis as those would have a direct conflict with the outcome variable, and would lead to an inaccurate model.

We opted for a linear regression model due to its capacity to quantify the marginal, and collective, effects of each predictor (CPI, housing supply, and interest rates) on housing affordability. This structure is well suited to match the topic and context of our analysis.

All modelling was conducted using the Base R package **(CITE HERE)**, specifically using the lm() function from the statistical package for linear regression analysis.

ALSO ADD A DAG HERE.

4 Results

Our results are summarized in Table 1.

Table Table 1 summarizes the performance metrics for our three models, with progressively added variables. Model 1, which only uses CPI as a predictor, achieves an R^2 of 0.370, indicating that CPI alone explains about 37% of the variance in the affordability index. Model 2 uses housing supply as an added predictor and achieves only a 0.005% increase in R^2, indicating that housing supply has a very minute impact on the variation of the affordability index. Moreover, model 2 only sees a very small reduction in the RMSE (850.31) showing a status-quo in predictive accuracy and model fit. This allows us to conclude that housing supply, when combined with CPI does not produce an impact that is statistically significant.

Model 3, which incorporates all three predictor variables, shows the largest improvement, with R^2 rising to 46.4% and RMSE dropping to 782.15, reflecting a stronger model fit. Moreover,

Table 1: Model Performance Summary showing improved fit and accuracy as Housing Supply and Interest Rates are added, with R^2 increasing from 0.370 in Model 1 to 0.471 in Model 3, and RMSE decreasing by over 70 points

	Model 1	Model 2	Model 3
(Intercept)	6113.14	6411.21	7499.80
	(469.64)	(518.18)	(507.91)
CPI (Index)	21.75	21.65	18.24
	(1.91)	(1.91)	(1.84)
House Supply (Month's Supply)		−41.41	−125.59
		(30.68)	(31.27)
Interest Rate (%)			198.23
			(31.41)
Num.Obs.	223	223	223
R2	0.370	0.375	0.471
R2 Adj.	0.367	0.370	0.464
AIC	3649.2	3649.4	3614.1
BIC	3659.4	3663.0	3631.2
Log.Lik.	−1821.611	−1820.691	−1802.059
F	129.840	66.073	65.101
RMSE	853.82	850.31	782.15

model 3 also sees the lowest values of AIC and BIC that indicate that model 3 is the best model as it balances fit and complexity most effectively. This progression highlights the benefit of adding contextual variables like housing supply and interest rates to better capture the accuracy of the model, as well as analyzing the marginal impacts of each predictor variable.

```
`modelsummary` 2.0.0 now uses `tinytable` as its default table-drawing
  backend. Learn more at: https://vincentarelbundock.github.io/tinytable/

Revert to `kableExtra` for one session:

  options(modelsummary_factory_default = 'kableExtra')
  options(modelsummary_factory_latex = 'kableExtra')
  options(modelsummary_factory_html = 'kableExtra')

Silence this message forever:

  config_modelsummary(startup_message = FALSE)
```
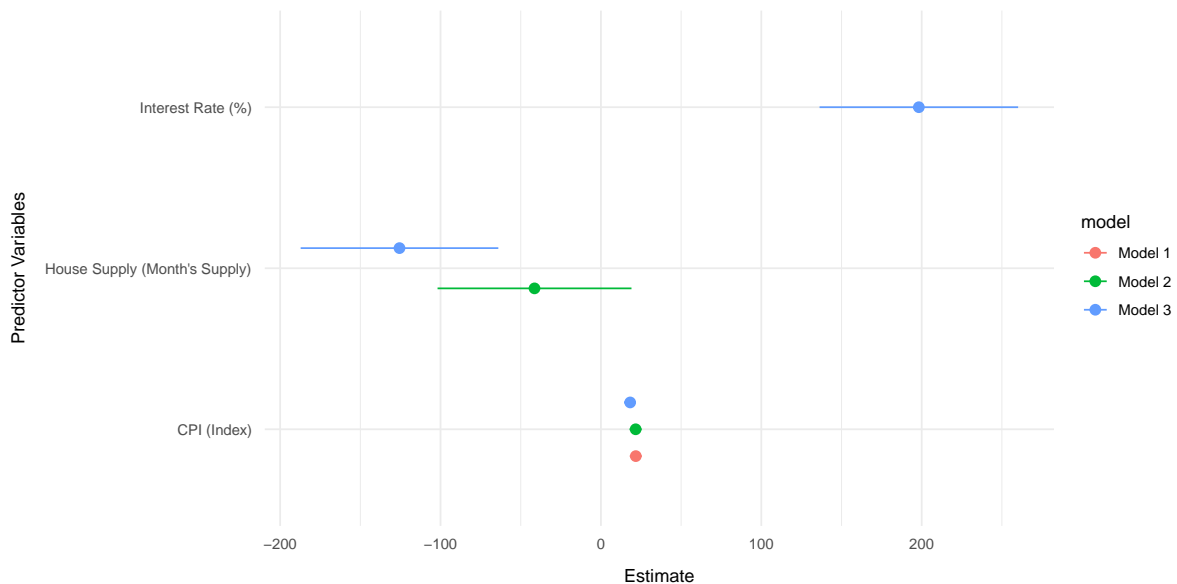

Figure 6: Coefficient estimates with 95% confidence intervals across three models, showing that CPI and interest rate positively affect the affordability index, reducing affordability, while an increase in housing supply improves affordability.

Moreover, Figure 6 visually represents the coefficients of the predictors in the three models, with their corresponding confidence intervals. Each point represents the estimated effect size

13

(coefficient) of a predictor, and the lines extending from each point indicate the range of uncertainty (confidence intervals).

The CPI coefficients remain positive for all three models, indicating that higher inflation (measured by CPI) is associated with a higher affordability index (i.e. less affordable housing). However, we see that the coefficients for housing supply are in the negatives for both models it's used as a predictor in. Suggesting that an increase in housing supply improves affordability, which is consistent with our expectations. Lastly, the coefficient for interest rate is strongly positive, indicating that higher interest rates significantly increase the affordability index (worsening affordability), likely due to higher mortgage payments. The wider confidence intervals of housing supply and interest rate depict uncertainty, while the absence of a confidence interval in CPI represents more statistically precise estiamtes. The results we have seen from Table 1 and Figure 6 are completely consistent with our broad economic expectations, however, this model has room for improvement in its predictive accuracy, model fit, and certainty.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point
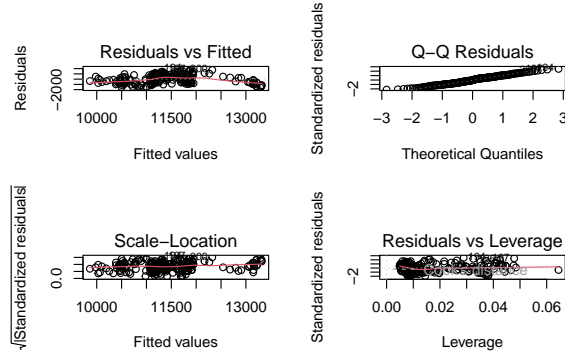
5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Diagnostics

(a) Trace plot

Figure 7: Diagnostic Plots of Model 3 show that assumptions are largely met

The diagnostic plot for Model 3 is given in Figure 7 and we notice the following:

1. **Residuals vs. Fitted**: The residuals display a somewhat random scatter around the zero line, indicating that the linearity assumption is largely met. However, slight deviations suggest potential minor non linearity, particularly in specific ranges of fitted values. This indicates room for improvement but does not severely undermine the model's reliability.

2. **Q-Q Plot**: The residuals follow a relatively straight line, indicating that the assumption of normality is generally satisfied. Small deviations at the tails are visible, but these are minor and unlikely to significantly impact inference or predictions.

3. **Scale-Location Plot**: The residuals appear to be evenly spread, demonstrating that the assumption of homoscedasticity (constant variance of errors) is broadly met. There are slight variations in spread, but these are relatively minor and unlikely to skew results.

4. **Residuals vs. Leverage**: A few points with slightly higher leverage are observed, but their Cook's distance values indicate that they are not excessively influential. These points warrant attention but do not dominate the overall model.

C References