Progress Report- Forecasting U.S Mortgage Rates: 2024 and Beyond

Group 59: Pallavi Singh, Themiya Dias Chandraratna, Abhi Patel, Roshni Patnayakuni, Nagendra

Git Hub Repository: https://github.gatech.edu/MGT-6203-Fall-2023-Canvas/Team-59

Scheller College of Business, Georgia Institute of Technology

MGT 6203: Data Analytics for Business

Dr. Jonathan Fan

November 4, 2023

Project Background

In the past 3 years, interest rates have more than doubled from roughly 3% for a 30-year fixed mortgage to well over 7%. October 2023 saw mortgage rates hit 7.79% which is the highest average 30-year mortgage rate since November 2000 (Rothstein, 2023). Mortgage rates are a key driver of the housing market and Central Bank rates are a prime factor affecting movements in mortgage rates (Sadorsky, 2015, p.822). Monetary policy changes affect market rates of interest because of federal fund rate pass-through (Hegwood & Tuttle, 2017, p.57). Although monetary policy changes affect market rates of interest, other factors can play a significant role (Arena et al., 2020, p.11). For example, Kliesen and Schmid (2004) found a statistically significant response of inflation expectations to surprises in the CPI, the core CPI, retail sales, and the NAPM index (p.10). Overall, the difference between inflation expectations and actual inflation tends to narrow (Kuncoro, 2020, p.76). This study aims to forecast U.S 30-year fixed mortgage rates based on macroeconomic data components.

Purpose

Interest rates have a significant effect on the economy, influencing consumer spending and manufacturing rates. Mortgage rates comprise a significant cost for housing buyers; a 4% increase in a 30-year fixed mortgage interest rate, from 3% to 7%, adds $280,740.62 to the cost of a $400,000 mortgage. Lower interest rates make homes more affordable, stimulating the housing market and encouraging new buyers. Existing homeowners benefit from refinancing at lower rates, saving thousands over the loan period. Lower mortgage rates increase disposable income, driving higher consumer spending on various goods and services, benefiting the broader economy. Higher rates deter potential buyers due to increased borrowing costs and may cool housing market activity. Rate fluctuations can cause financial stress for homeowners, especially those with adjustable-rate mortgages. Accurate mortgage rate forecasts are valuable for investors, firms, government entities, and homebuyers. The models and techniques from this study could potentially provide accurate mortgage rate forecasts and trends for 2024 and beyond.

Data Preparation

An in-depth analysis of Federal Reserve Economic Data (FRED) was performed, focusing on key U.S. economic categories: Money, Banking, & Finance, Population, Employment, & Labor Markets, National Accounts, and Production & Business Activity. Discontinued data sources were avoided, and calculations reflecting current best practices were prioritized. Selected variables included Consumer Price Index (CPI) for all items, Total Federal Public Debt, Households Net Worth, Housing Inventory Active Listing Count, Industrial Production Consumer Goods, Industrial Production Total Index, 3-Month Interest Rates, 10-Year Treasury Yield, Job Openings Total in non-farming, M2 money supply, Producer Price Index (PPI) for all commodities, Real Gross Domestic Product (GDP), Unemployment Rate, and Velocity of M2 money supply.

Datasets were gathered from FRED and stored in a MySQL relational database on Amazon Web Services (AWS) Relational Database Service (RDS). Data was manipulated to convert record dates to Year-Month format (%Y-%m) for consistent matching. Initially, joining all selected factors by record date yielded only 28 records for 14 predictors, falling short of the recommended 10 records per predictor variable for regression analysis. Further joins were explored, leading to a dataset comprising 8 predictors (Industrial Production Consumer Goods, CPI, Households Net Worth, 3-Month Interest Rates, Job Openings Total in non-farming, M2 money supply, and Velocity of M2 money) with the response variable of 30-Year Fixed Rate Mortgage Average. To handle multiple records per month for mortgage rates and money supply, average monthly values were used after additional data manipulation. After conducting exploratory data analysis to assess the reliability of the selected predictors, a Java code was developed to calculate period lag on mortgage rates, enabling the computation of forecast horizons spanning 3, 6, and 12 months. This step enhanced the predictive capability of the dataset.

Future Scope for Data Preparation

Imputation

In the data preparation process and leveraging inner joins, the dataset was cleared of null values and the time-period and dataset were created such that any factors or fields that proved to contain nulls were avoided. Imputation can be leveraged to fill null values and regression can be used on predictor variables to provide accurate values.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial initial phase following data collection and pre-processing. During this stage, the data is visualized and manipulated without any underlying assumptions. This process aids in assessing the data's quality and serves as a foundation for model building. To examine correlations between attributes, a correlation matrix was plotted, revealing the correlation coefficients between the variables.

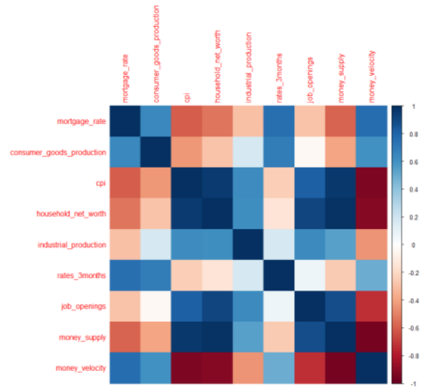


Fig.1. Correlation Matrix Heatmap

From the correlation matrix above [Fig.1], high correlation between .8 and 1 is observed between velocity with money_supply, cpi, and household_net_worth. A threshold of 0.7 can be used for feature selection.

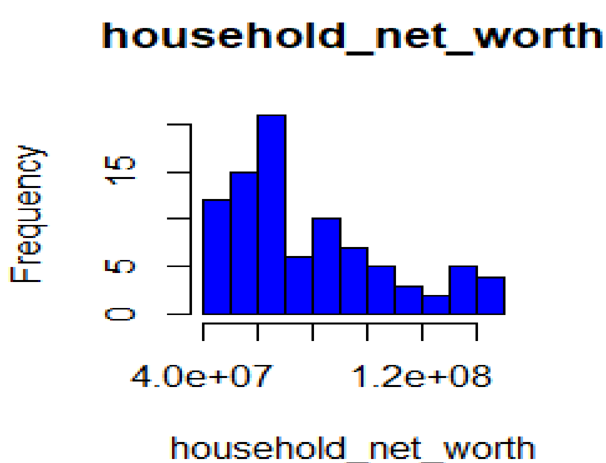


Fig.2. Histogram of Household Net Worth

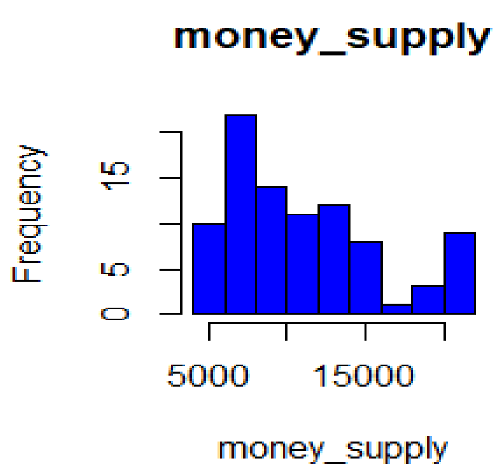


Fig.3. Histogram of Money Supply

Histograms are among the most useful EDA techniques, and allow you to gain insights into your data, including distribution, central tendency, spread, modality and outliers (Komorowski et al., 2016). We created histograms for all variables, revealing skewedness in most cases (all plots not shown here). These variables tend to deviate from a normal distribution. Logarithmic transformation can be applied to address this skewedness.

The below figure shows pair plots for the mortgage rate and all the other numeric variables. From the graphs, we can see that variables like CPI and household net worth are directly correlated. Variables like mortgage rate, industrial production, and rates_3 months are scattered randomly and hence not correlated. We can see similar observations for CPI vs household net worth and mortgage rate vs industrial production.

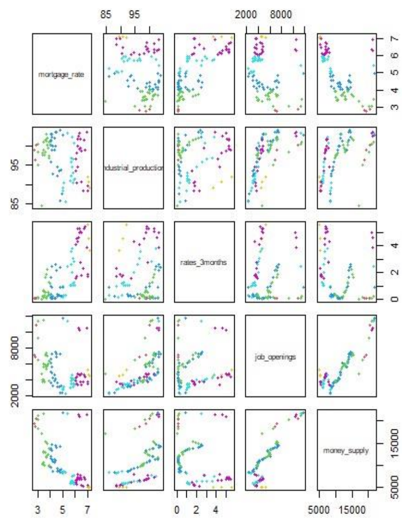
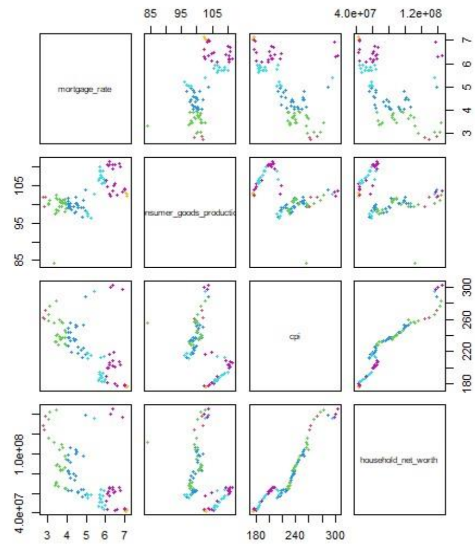


Fig.4. Pair Plot of Mortgage Rate and Predictors

Fig.5. Pair Plot of Mortgage Rate and Predictors



Future Scope for EDA

Feature Selection

Feature selection is the process of reducing variables to develop a simple and efficient model. It is an important step to improve the cost of computation. Many models, especially those based on regression slopes and intercepts, will estimate parameters for every term in the model. Because of this, the presence of non-informative variables can add uncertainty to the predictions and reduce the overall effectiveness of the model (Brownlee, 2019).

1) Pearson coefficient - Pearson's Method is one of the most common techniques used for a linear correlation with numerical variable input and numerical variable output.

2) VIF (Variance Inflation factor)- VIF is used to check multicollinearity among independent variables in regression. High VIF leads to multicollinearity which means input is not independent and influences each other.

Analytical Approach

```
Call:
lm(formula = mortgage_rate ~ ., data = corr_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.84352 -0.31655  0.00891  0.29326  1.17226 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)             -7.079e+00  3.362e+00  -2.105   0.0383 *  
consumer_goods_production -4.494e-03  3.035e-02  -0.148   0.8826    
cpi                      7.465e-03  1.157e-02   0.645   0.5206    
household_net_worth     -7.407e-08  1.778e-08  -4.166 7.71e-05 ***
industrial_production   -4.978e-02  2.749e-02  -1.811   0.0738 .  
rates_3months            2.733e-01  6.095e-02   4.484 2.39e-05 ***
job_openings             1.118e-04  9.968e-05   1.122   0.2652    
money_supply             6.790e-04  2.185e-04   3.107   0.0026 ** 
money_velocity           7.813e+00  1.865e+00   4.190 7.06e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4391 on 81 degrees of freedom
Multiple R-squared:  0.886,	Adjusted R-squared:  0.8748 
F-statistic: 78.72 on 8 and 81 DF,  p-value: < 2.2e-16
```

Fig. 6 Initial Regression Model

```
> summary(model_3_months)

Call:
lm(formula = X3monthForecast ~ consumer_goods_production + cpi + 
    household_net_worth + industrial_production + rates_3months + 
    job_openings + money_supply + money_velocity, data = file_path)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.92570 -0.26591  0.01864  0.30567  1.06745 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -1.371e+01  3.653e+00  -3.752 0.000329 ***
consumer_goods_production  3.361e-03  3.297e-02   0.102 0.919055    
cpi                        1.121e-02  1.257e-02   0.892 0.375255    
household_net_worth       -6.553e-08  1.932e-08  -3.391 0.001077 ** 
industrial_production     -5.521e-02  2.987e-02  -1.849 0.068154 .  
rates_3months              1.131e-01  6.623e-02   1.707 0.091677 .  
job_openings               1.738e-04  1.083e-04   1.605 0.112451    
money_supply               7.327e-04  2.374e-04   3.086 0.002774 ** 
money_velocity             1.035e+01  2.026e+00   5.108 2.12e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4771 on 81 degrees of freedom
Multiple R-squared:  0.8647,	Adjusted R-squared:  0.8513 
F-statistic: 64.69 on 8 and 81 DF,  p-value: < 2.2e-16
```

Fig.7. Linear Regression Model for 3month Forecast

Initially, a regression model was trained on mortgage rates and 8 selected predictors to assess reliability. The model showed high significance for several predictors, resulting in an adjusted R^2 value of 0.8748, signifying a good fit. Following Sadorsky's (2015) approach, lag periods of 3, 6, and 12 months were used for forecasting mortgage rates. The 1-month period was omitted due to its limited relevance in the context of financial quarters (p. 823).

```
Call:
lm(formula = X6monthForecast ~ consumer_goods_production + cpi + 
    household_net_worth + industrial_production + rates_3months + 
    job_openings + money_supply + money_velocity, data = file_path)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.24870 -0.34352  0.04104  0.32007  1.24992 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -1.429e+01  3.982e+00  -3.588 0.000569 ***
consumer_goods_production  3.552e-02  3.594e-02   0.988 0.325858    
cpi                        2.200e-02  1.370e-02   1.606 0.112273    
household_net_worth       -4.938e-08  2.106e-08  -2.345 0.021466 *  
industrial_production     -9.416e-02  3.255e-02  -2.893 0.004902 ** 
rates_3months              3.215e-02  7.218e-02   0.445 0.657195    
job_openings               3.279e-04  1.181e-04   2.778 0.006794 ** 
money_supply               5.221e-04  2.587e-04   2.018 0.046923 *  
money_velocity             9.746e+00  2.208e+00   4.414 3.11e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.52 on 81 degrees of freedom
Multiple R-squared:  0.8418,	Adjusted R-squared:  0.8261 
F-statistic: 53.86 on 8 and 81 DF,  p-value: < 2.2e-16

> 
```

Fig.8. Linear Regression Model for 6month Forecast

```
> summary(model_12_months)

Call:
lm(formula = X12monthForecast ~ consumer_goods_production + cpi + 
    household_net_worth + industrial_production + rates_3months + 
    job_openings + money_supply + money_velocity, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.57349 -0.32744  0.01377  0.37104  1.27092 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)   
(Intercept)               -1.131e+01  4.529e+00  -2.498  0.01457 * 
consumer_goods_production  6.207e-02  4.032e-02   1.539  0.12773   
cpi                       -3.454e-03  1.611e-02  -0.214  0.83073   
household_net_worth       -6.061e-08  2.406e-08  -2.519  0.01380 * 
industrial_production     -8.487e-02  3.777e-02  -2.247  0.02744 * 
rates_3months             -1.433e-04  8.119e-02  -0.002  0.99860   
job_openings               3.529e-04  1.322e-04   2.669  0.00923 **
money_supply               7.087e-04  2.949e-04   2.403  0.01861 * 
money_velocity             8.489e+00  2.489e+00   3.411  0.00102 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5803 on 79 degrees of freedom
Multiple R-squared:  0.796,	Adjusted R-squared:  0.7754 
F-statistic: 38.54 on 8 and 79 DF,  p-value: < 2.2e-16

> 
```

Fig.9. Linear Regression Model for 12month Forecast

The 3-month regression model (Fig. 6) demonstrates a significant relationship between household_net_worth, rates_3months, and money_velocity with mortgage rates at a 99.9% confidence level. In the 6-month forecast model (Fig. 8), money_velocity remains significant at 99.9% confidence, and job_openings and industrial_production also exhibit significance. Similar patterns emerge in the 12-

month forecast model (Fig. 9). These significant parameters can inform further analysis for creating a new model. However, with the increase in the forecasting period, the adjusted R^2 decreases, indicating reduced reliability of the independent variables in predicting mortgage rates. Testing through cross-validation can assess the accuracy of the models and determine the best one. Based on the fit using adjusted R^2, the regression model for 3-month forecasts proves to be the most accurate.

Future Scope for Analytical Approach

ARIMAX

Sadorsky (2015) identified ARIMAX and OLS as the most effective methods for forecasting 6 and 12 months ahead (p.824). Since we have already examined the OLS approach, exploring ARIMAX for model comparison could be beneficial.

Engle-Granger test

Ordinary Least Squares regression might yield misleading outcomes in time-series analysis (Hegwood & Tuttle, 2017). The Engle-Granger test can help exclude short-run dynamics.

Deep Learning

Albanesi and Vamossy (2019) emphasize the need for deep learning to capture the complexity of consumer default behavior, asserting that all deep learning models significantly outperform logistic regression (p.3). Given our use of similar high-dimensional data with intricate interaction patterns, standard regression might yield poor results when compared in cross-validation (Albanesi and Vamossy, 2019, p.1).

Future & Adjustments

Sadorsky (2015) mentions an 8-year window of monthly data is long enough time to cover most business cycles and include enough observations to ensure good parameter fit for the forecasting methods (p.822). We can leverage this observation in our testing to analyze the results of our model. Furthermore, imputation can be used to potentially enhance data quality. Further regression analysis and running models like Decision Tree on specific predictors can potentially enhance the relevance and significance of those predictors and the model at large.

GitHub Repository

https://github.gatech.edu/MGT-6203-Fall-2023-Canvas/Team-59

Works Cited Section

Albanesi, Stefania, and Domonkos F. Vamossy. 2019. “Predicting Consumer Default: A Deep Learning Approach.” National Bureau of Economic Research, working paper no. w26165, August. Available at https://doi.org/10.3386/ w26165

Arena, M., Bella, G., Cuevas, A., Gracia, B., Nguyen, V., & Pienkowski, A. (2020). It is Only Natural: Europe’s Low Interest Rates (Trajectory and Drivers). *IMF Working Papers*, *2020*(116), 1–59. https://doi.org/10.5089/9781513549170.001

Brownlee, J. (2019, November 26). *How to Choose a Feature Selection Method For Machine Learning*. Machine Learning Mastery. https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/

Hegwood, N., & Tuttle, M. H. (2017). CONVENTIONAL MORTGAGE INTEREST RATE AND THE EFFECTIVE FEDERAL FUNDS RATE PASS-THROUGH. *Journal of Business Strategies*, *34*(1), 57–74. https://doi.org/10.54155/jbs.34.1.57-74

Kliesen, K., & Schmid, F. (2004). Monetary Policy Actions, Macroeconomic Data Releases, and Inflation Expectations. *Federal Reserve Bank of St. Louis Review*, *86*(3), 9–21. https://doi.org/10.3886/ICPSR01301

Komorowski, M., Marshall, D.C., Salciccioli, J.D., Crutain, Y. (2016). Exploratory Data Analysis. In: Secondary Analysis of Electronic Health Records. Springer, Cham. https://doi.org/10.1007/978-3-319-43742-2_15

Kuncoro, H. (2020). Central Bank Communication and Policy Interest Rate. *International Journal of Financial Research*, *12*(1), 76–91. https://doi.org/10.5430/ijfr.v12n1p76

Rothstein, R. (2023, November 3). *Mortgage Rate Forecast For 2023* (C. Jennings, Ed.). Forbes Advisor; Forbes. https://www.forbes.com/advisor/mortgages/mortgage-interest-rates-forecast/

Sadorsky, P. (2015). Forecasting Canadian mortgage rates. *Applied Economics Letters*, *23*(11), 822–825. https://doi.org/10.1080/13504851.2015.1111981