

## Statistical Analysis of Economic Variables and the Mortgage Rate in the United States

Purpose: Use various sources (suggestion: most likely you can find on FRED - Federal Reserves Economic Data - database) to get data series of mortgage rate and other variables, and analyze the collinearity among the independent variables, as well as applying Non-Linear functional forms.

Task:

1) Get the following monthly data series, starting from 30 years ago (code in brackets), line them up with the same starting points to endpoints:

Y = 30-year mortgage rate (mortg) or any similar data

X1 = federal funds rate (fedfunds)

X2 = inflation rate (cpi)

X3 = nominal gdp (ngdp); GDP is usually reported quarterly, so make it monthly by extrapolation.

X4 = 3-month Treasury Bills (tb3ms)

X5 = 20-year Treasury Bonds maturity rate (gs20)

X6 = House price index (choose what you think is relevant)

X7 = ..

X8 = ..

2) Add two more variables of your own choice; look for your own sources, you can use binary (and then create Interactive on some variables - just a suggestion); improvise as you need to, in terms of duration and frequency of your data; but I suggest to start data from 1990 until the most current to ensure you get large enough samples to satisfy normality assumption, as well as seeing the up and down of the economy since that year, and how it changes the variables.

3) For each explanatory variables, indicate the effect (positive or negative) that you anticipate, before running a regression, to start you thinking of the logic of your model.

4) Estimate OLS with all the variables. Based on the output, check for t-test and F-test at 5% significance level, and sum up the results.

- 5) Check for the correlation coefficients of the x-variables, assess if there is a case of multi-collinearity; identify variables that are insignificant that you can drop.
- 6) Analyze the residuals against each of the x-variables using textplot or scatter. What do they say?
- 7) Change one or two variables into different format (logs, squares, etc.); if there's dummy vars, create interactive terms; then include them into your model.
- 8) Carry out HSK test, make conclusion; if there's a case of HSK, carry out WLS (choose what variable as the weight), obtain the estimates. Eliminate some variables as necessary. Modify your model accordingly. Check t, F, p-values for significance.
- 9) As you start with the original model, checked for significance of x-variables, omit one at a time as necessary, each time re-estimate and justify each omission; Use the stargazer commands (in R) to gather your regression outputs later, giving you a more professional output comparison.
- 10) What is your final model? Do the signs or the (non)linearity in the final model agree with your intuition?
- 11) Write up a cover page for your assessment and conclusion, and attach your summary of outputs from stargazer; you can put the other regression outputs on Appendix at the very last page. Make it organized and presentable. Submit your report here on Canvas.

====

Additional notes about the Data from FRED;

There are many variations of both CPI and GDP. Each has its usefulness. And it depends on the researcher on how to use it, usually choosing it based on what you need. In a paper, usually we describe the data used, along with reasons why we are using it.

For CPI: Since we are looking at the long-term mortgage rate, I would suggest to use Urban (since mortgages in cities are more prevalent), and choose the one that exclude (less) Food and Energy (because prices of food and energy fluctuates more and usually - and hopefully - temporary in the short run).

As for GDP, I suggest to choose nominal not Real (since we already have CPI variable; if we use Real GDP, there will be correlation issue with CPI). Also, it's better if you can use Seasonally-Adjusted data (taking out the outliers from holiday fluctuations). Since the GDP is reported quarterly, to get the monthly trend, do extrapolation. You can lookup this source, scroll down for Extrapolation for Non-Linear Data using Trendline:  
<https://www.exceldemy.com/extrapolate-data-in-excel/>