Bursting Bubbles

Benjamin Kilian and Toluwase Afolayan

March 2020

Abstract

Inspired by the Housing Crisis of 2008 and other similar recessions, we have selected data related to the economy at large and more specifically the housing market in order to analyze the behavior of these data and their relationships. Specifically, we looked at how overall economic activity and housing loan data may effect inequality in American society, as measured by discrepancies in the mean and median income. By means of linear regression, we found that there are relationships between these values, with an especially distinct linear relationship between the growth of the American economy and the inequality within it.

1 Background

The American economy is the largest national economy in the world, but contains high degrees of inequality despite its success on the world stage. To explore this further, we looked at several quantities and their respective behaviors that could demonstrate these concepts. The strength of the U.S. economy is usually defined by the strength and success of markets. In 1999 the U.S. government deregulated banks in such a way that allowed them to speculate in a manner similar to the way hedge funds operate, prompting them to start seeking ways to outperform the market, and make a higher return on new investments. Banks also spent millions lobbying state legislatures to relax laws. Most of these laws centered around limiting the risk that borrowers can have, when taking out loans. The financial success of bankers selling these mortgage-backed securities, caused their demand to skyrocket leading to a lowering of interest rates and borrowing requirements and standards. The purpose of a mortgage-backed security allowed lenders the ability to bundle loans and resell them. Because of the introduction of interest-only loans, the risk of a loan, and therefore a MBS could be passed on to the buyer of the security. This was only low risk so long as the housing market continued to rise. The federal government started raising interest rates, and the housing market fell. This resulted in mass defaults amongst subprime mortgage holders. At this point all these large banks that had been dealing in these mortgage backed securities found themselves in possession of millions of dollars of worthless investments. This sudden downturn in the economy though it effected the economy as a whole, and most big banks to some extent, at the end of the day, it was the middle and lower class that bared the brunt of the recession while the Obama Administration bailed out the big banks that had created the bubble in the first place.

The M2, or money stock, is simply the total amount of money in the American economy. The Monetary Base, which we denote MB several times during our report, is the amount of money which circulates in a given week; we adapted this to monthly values in this case. To account for the fact that more total money in the economy may naturally cause more money to circulate, we also include the ratio of the monetary base to the money stock, $\frac{MB}{M2}$. As during the 2008 housing crisis, loans were a highly important part of the reason for the economic recession that followed. Therefore, we also analyze the total sum of all real-estate loans in the United States. Finally, the Case-Shiller index is an index which measures the prices of a home when it is sold and re-sold at different times. This index is used to determine the strength of the housing market and the overvaluation of homes. In order to evaluate how the economy was performing, we looked at both the unemployment rate and at the difference in mean and median incomes in American society.

2 Methods

We obtained data from FRED, the Federal Reserve Economic Database. Most of our data was reported monthly, however some were annual, weekly or daily. Since the majority was monthly data, we adapted all data to monthly. For weekly and daily data, we simply used the first data point reported of each month, whereas for annual data we used the value of each year for each of the 12 months in that year.

After scraping data, we first checked to see if the input data had linear relationships with what we suspected to be output data. Namely, we determined that things like loans being given, the Case-Shiller Index, and the total money being spent in the economy act as reasonable inputs which can give rise to economic trends like unemployment or inequality. To analyze the data, we ran a linear regression to obtain a linear model on training data, which was a randomly selected are subset of 75 percent of our obtained data. The remaining 25 percent of the data was used as the test set on which we analyzed the performance of our model. Our model will give us a predicted output y_i for some test data-point $x_{1_i} + x_{2_i} + ... + ...x_{p-1_i}$as follows:

$$y_i = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + ... + ...\beta_{p-1} x_{p-1_i} + \epsilon_i,$$

where ϵ_i In order to test this performance, we used the F-Statistic, and tested the null hypothesis, that none of our inputs had any relationship with our output, i.e.

$$H_0:$$
$$\beta_0 = \beta_1 = ... = \beta_{p-1} = 0,$$

The One-Way ANOVA Table (using shortcut formulas)

Source of Variance	Sum of Squares	Degrees of freedom (df)	Mean Squares (SS/df)	F-Ratio (F-Statistic)
(Between) Group or Treatment	$SSTr = \sum n_i \bar{x}_i^2 - N\bar{\bar{x}}^2$	k-1	$MSTr = SSTr/(k-1)$	$F = MSTr/MSE$
(Within Group) or Error	$SSE = SSTo - SSTr$	N-k	$MSE = SSE/(N-k)$	
Total	$SSTo = \sum x^2 - N\bar{\bar{x}}^2$	N-1		

Figure 1: ANOVA table (Source: Dr. Joseph Ortiz, Kent State University

where each β is given to us by our magical linear regression machine, Python's sklearn library and p is our number of parameters. The F-Statistic is given to us by using an ANOVA table, as seen is Figure 1. To find the F-Statistic of our model, we calculate the sum of the squares of each error, as well as the sum of the squares given by the regression itself, i.e. the sum of the squares of the difference between each point and the mean value.

2.1 Unemployment as output

To make a model of this data, we had to find a suitable output for measuring the performance of the economy; or at least the portion of the economy that could be effected by our chosen inputs. We began by looking at the unemployment rate, but when we looked at the plots of each of our input variables vs. national unemployment, as seen in figure 2, we do not see any linear relationships. While one can observe some interesting trends in the data, they are not linear trends, and therefore we choose not to proceed with a linear regression. Instead we chose to look for new quantities that may have linear relationships with our input quantities.

3 Results and Analysis

For our analysis, we narrowed our inputs to the 5 discussed in the previous section:

1. The M2 index: The total money in the economy.
2. The monetary base: The amount of money circulated in a month.
3. The ratio of these two values: the amount of money circulated in the economy divided by the total money.
4. The total sum of all real-estate loans in the US.
5. The Case-Shiller index.

3.1 Measuring Inequality

After we saw that our parameters did not have a linear relationship with unemployment, we decided to look at inequality in the economy. First we had to determine how to measure inequality in the United States, which we ultimately decided to measure using the income distribution. Namely, the difference in the mean and median income in the United States. As an economy becomes more unequal, the wealthier people will amass more money and will bring up the mean income,

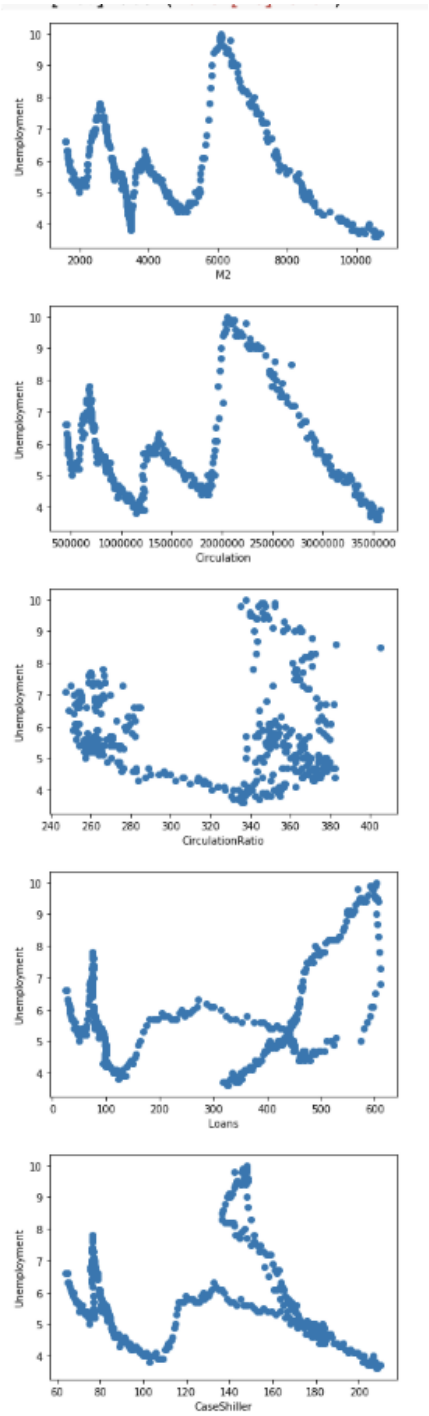

Figure 2: Various economic indicators vs. national unemployment rates

while the median income will not increase so much. The bigger this gap, the more unequal we will say that the society is. Running the same regression as before but with the difference in mean and median income as the output, we observed much better results based on an initial eye-test, as seen in Figure (N).

3.2 Results of Linear Regression

After running our linear regression using Python's sklearn library, we obtain the following coefficients, with the first element, β_0 being the intercept of our model and each of the following elements corresponding to the parameters that we input.

$$\beta = \begin{bmatrix} -75272 \\ 14.01 \\ -.025 \\ 168.71 \\ 7.97 \\ -.37 \end{bmatrix} \tag{1}$$

Therefore, our model predicts:

$$\mathrm{I}_{mean_i} - I_{med_i} = \begin{bmatrix} -75272 \\ 14.01 \\ -.025 \\ 168.71 \\ 7.97 \\ -.37 \end{bmatrix} \begin{bmatrix} 1 M2_i \\ MB_i \\ \frac{MB_i}{M2_i} \\ L_i \\ CS_i \end{bmatrix} + \epsilon_i$$

3.3 Comparing Model to Sub-Models

After we fit our model and got our coefficients, we tested how our model compares to each of its sub-models using the F-Test. Since we had only 5 inputs and a test dataset of only 92 samples, it was not too computationally expensive and we were able to simply test each possible subset.

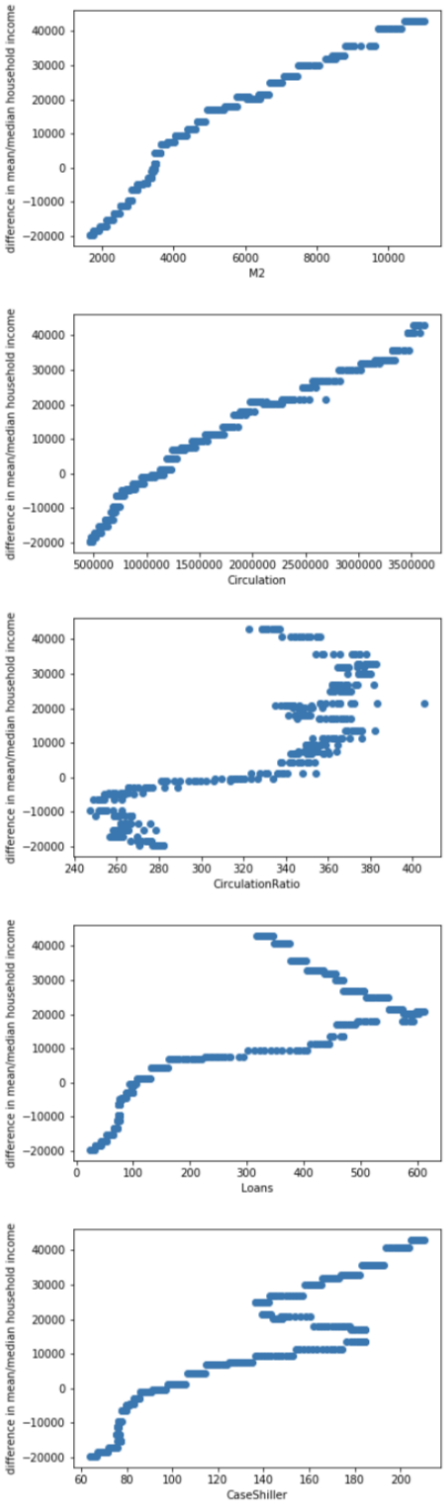

Figure 3: Various economic indicators vs. Inequality

F-Test Results				
Inputs	P	F-Statistic	Req. to Reject (for α=.05)	Reject?
1	1	223.98	3.94	YES
2		90.49		YES
3		88.37		YES
4		92.93		YES
5		92.55		YES
1,2	2	47.88	3.09	YES
1,3		69.80		YES
1,4		118.96		YES
1,5		110.57		YES
2,3		41.12		YES
2,4		44.79		YES
2,5		44.78		YES
3,4		45.62		YES
3,5		43.71		YES
4,5		45.98		YES
1,2,3	3	789.91	2.70	YES
1,2,4		32.03		YES
1,2,5		31.58		YES
1,3,4		45.07		YES
1,3,5		46.08		YES
1,4,5		78.31		YES
2,3,4		27.07		YES
2,3,5		27.13		YES
2,4,5		29.55		YES
3,4,5		30.07		YES
1,2,3,4	4	2192.49	2.46	YES
1,2,3,5		563.35		YES
1,2,4,5		23.76		YES
1,3,4,5		33.47		YES
2,3,4,5		20.09		YES
All	5	1753.07	2.31	YES
Input Keys 1- M2 2- Monetary Base 3- M2/Monetary Base 4- Total US Loans 5- Case-Shiller Index				

Looking at the F-statistics, we can reject the null hypothesis in every case. For each variable, the F-statistic is higher than the threshold, which we found in a table on Stanford's website.

4 Conclusion

Our F-Statistic data demonstrates clear correlation between the each of our chosen parameters and inequality in the United States, but the most clear relationships exist between inequality with economic growth and activity. This suggests that as more wealth comes into the American economy it ends up moving towards those who already have wealth. To put it simply, the rich get richer.

Going forward, it would be interesting to analyze similar trends in other countries, or in the world as a whole. Specifically, one could try to find a relationship between forms of government and these same metrics related to growth and inequality. Would we observe similar trends in a country with more social safety nets like Norway, or in a country with a more authoritarian form of capitalism like China? What would this trend look like analyzed over the world as a whole?