Midterm Project

Alison Pedraza MA678

12/11/2020

What is the Relationship Between the Wealth of the 1% and the US Stock Market?

Abstract

In this study, I looked at what effect the net worth of the top 1% of income earners as well as their equity could have on the market cap of the U.S. stock market. Since this is a causal question (Is the stock market currently over valued because there is an influx of money coming from the top 1% of income earners?), I used a propensity score model with a linear regression outcome model. However, I also used a multilevel linear regression model since I was not satisfied with the former model's results and was not convinced that, although this seems like a causal question, a causal approach was the logical approach to take.

Summary

Over the past 40 years there has been a decrease in wage growth of the U.S. labor force while the net worth or yearly income of the top 1% of income earners has increased steadily every year. At the same time, the value of stocks in the U.S. stock market has increased without logical reasoning. In other words, the total value of shares in the stock market is larger than the gross domestic product (the value of what those companies have produced). For this reason, the stock market is considered to be overvalued. Why is there so much money invested into the stock market when the total value of all finished goods and services of the United States is much lower? What I am asking is: has the increase in the net worth of the 1% income earners caused the overvaluation of stocks? Is the money that is not going to the labor force ending up in the stock market through the 1% earners of the United States?

For this study, I ran two models: one using a propensity score model, and another using a multilevel regression model.

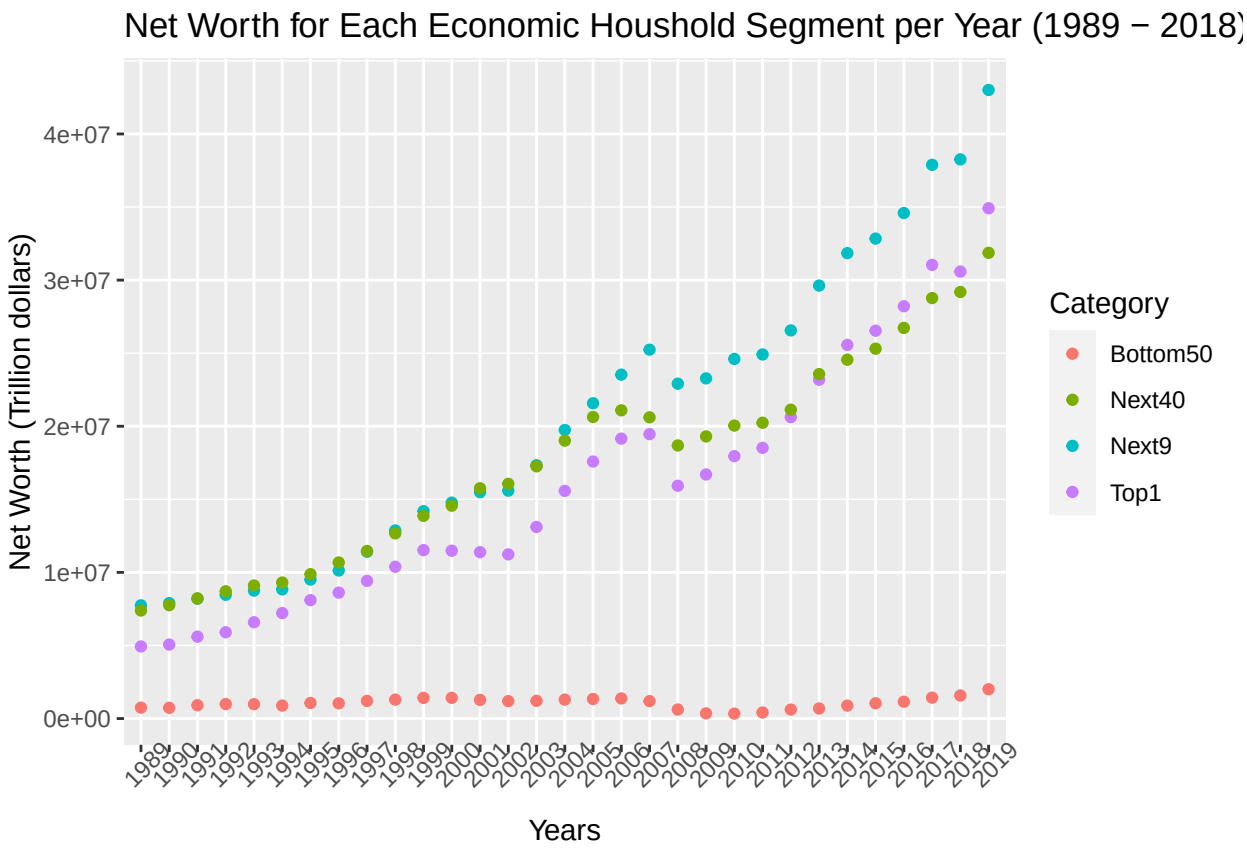
My predictors are the following:

- Market Capitalization
- Net Worth of all the income earners (Top1, Next9, Next40, Bottom50)
- Corporate Equities and Mutual Funds held by the above four categories
- Labor share (income earned for every dollar produced through labor)
- Output (the amount of goods and services produced in (dollars))

EDA Plots

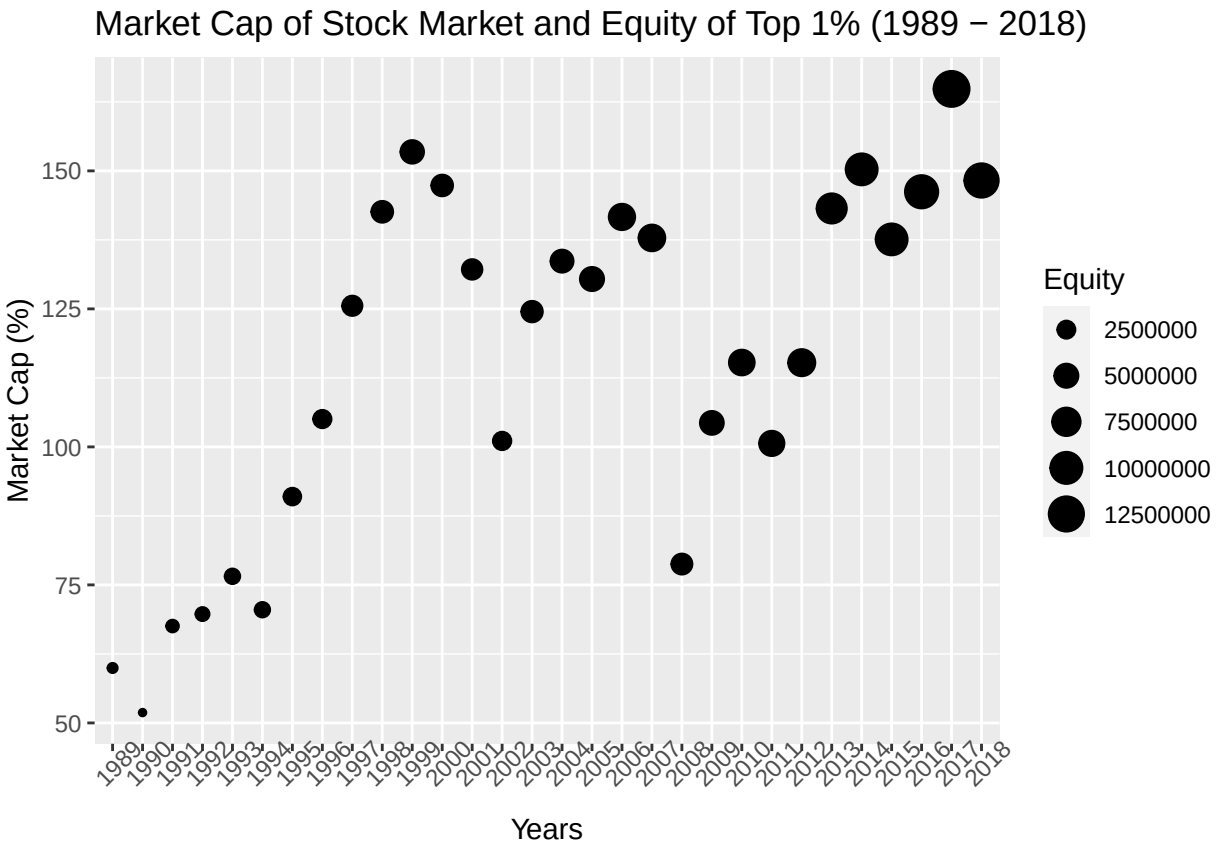
Some EDA plots were done to see if infact there is a difference in net worth and equity between the different income earning households and to see if there was a trend of increased value of the market cap with time. I wanted to see if there could be a positive relationship between the net worth and equity of the top 1% with the market cap. A positive realtionship could suggest that there is some causal relationship.

Net Worth for Income Levels per Year If we look at equity holdings and net worth for each of the income levels, we see that while the net worth and equity holdings have remained the same for the bottom 50% of income earners, they have increased for both the top 9% and top 1%. If market cap has also increased during this same time frame it could suggest that there is a relationship between the predictors and the market

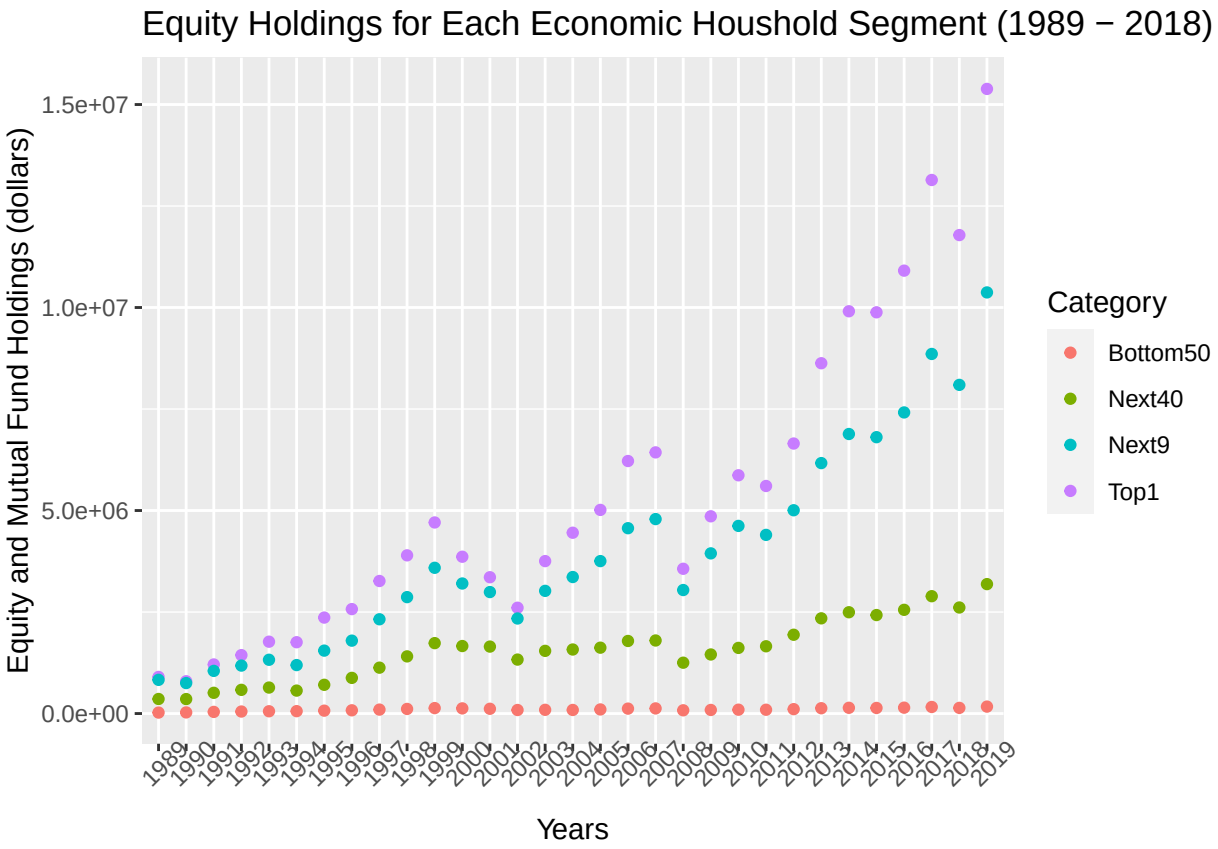


cap.

Market Cap and Equity of the Top 1% per Year From the below graph, we can see that indeed as market cap has increased every year along with the equity of the top 1% of income earners.



Corporate Equities Holdings for Each Economic Category vs Market Cap Below is a chart of corporate equities held by each of the income earner levels. As can be seen, the equity held by the top 1% has increased steadily and dramatically since 1989 while that cannot be said by the next 90% of economic income earner levels.



My Assumptions

My assumptions are that the net worth of the 1%, their corporate equity and mutual fund holdings, the labor share of the US worker, and the output produced all have some relationship with the increase in market capitalization of the U.S. stock market. Below is a diagram showing my assumptions. Since 1989, as output has increased, the labor share for the US worker has steadily decreased while maintaining a steady average weekly hours rate. Where has the difference in income gone if not to the workers who have produced it?

The increase in dollar output produced is not going to the majority of the American work force. The net worth of the top 1% is increasing as well as their equity. Therefore, my assumptions are that increased output, reduced labor share by the majority of the US labor force, the difference in equity held by the four categories of households, and the net worth of the 1% is causing the stock market to be inflated.:

Increased output, labor share, and equity are affecting the top 1% of earners who are then investing those dollars into the stock market. Therefore, output, labor share, and equity are also affecting the market capitalization.

A diagram of my assumptions can be found in the Appendix.

Causal Inference

Causal inference was used to see if the top 1% income earners have an effect on the market cap. A propensity model was fitted in which the three confounders were labor_share, equity, and output. My groups were Top1%, Next9%, Next 40%, and Bottom 50%. My counterfactual argument is: what if all of the income

earners had the same income (or more realistically, all had income levels that were very close), would the market cap still be affected? Below is a histogram of the propensity scores where:

- Category 1 = Top 1%
- Category 2 = Next 9%
- Category 3 = Next 40%
- Category 4 = Bottom 50%

The histogram showed a higher count for the Top 1% and much fewer counts for the Bottom 50%. This may be because of their value in term of net worth. The x-axis of this histogram should represent the probability of being in one of the four categories. In this case, there is less probability to be in the top 1% than it is to be in the Next 40% and Bottom 50%.

The histogram can be found in the Appendix.

Fitting the Outcome Model After doing the propensity scores, I weighted the four categories. Since the data was done over time, it looks like the weighting considered each year that came up for all four categories. In this sense, in a way, it considered each year to have a new "individual" of that category.

For the outcome model I used a basic linear model : lm()

model <- lm(Market_Cap ~ Category, data = df2, weights = w_ate)%>% tidy()

I used this model since my outcome (Market Cap) is continuous and from the EDA there seems to be a linear relationship between the outcome and the predictors. A summary results of this model is below:

Although the p-value is less than 0.05, I am not convinced that this was the right approach to this question. Although this was a causal question, there was no treatment involved. It could be thought that the treatment was the degree of net worth (high net worth or very low net worth). Or, the argument could be that the market cap of the stock market would increase regardless of which category held the most weatlth (if the Top 1% category held the least amount of wealth and the Bottom 50% held the most wealth).

```
##-----Fitting Outcome model
df2 <- df[-c(113),]
model <- lm(Market_Cap ~ Category, data = df2, weights = w_ate)%>% tidy()

#summary(model)
print(model)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   134.        8.51     15.8  1.06e-30
## 2 Category       -6.23      2.86     -2.18 3.14e- 2
```

Multilevel model:

Since I was not satisfied with the result from the previous model, I tried to answer the question using a multilevel model linear regression model.

In my multilevel model I used stan_lmer to see if I could get a better result. I included the same predictors : equity, labor.share, and net worth with market cap as the output.

The results of my model show the coefficient for net worth is zero while the coefficient for labor share had a number of different slopes. This tells me that as the market cap increases, the net worth of the 1% does not change. This is not what was seen in the EDA.

The results also show that the slopes for all the labor share is positive. However, even though they are positive, the rate of the slopes' decreases as the labor share decreases. This was visible in the EDA.

The residuals for my multilevel model did not show a pattern and most of the residuals were very close to zero. This seems to say that the observed values are not different from the predicted values. However, since the model gave me a coefficient of zero for net worth, these residuals may simply be reflecting that result.

```
model2 = stan_lmer(Market_Cap ~ Corporate.equities.and.mutual.fund.shares + Labor.share +Net.worth + Ou
```

```
## stan_lmer
##  family:       gaussian [identity]
##  formula:      Market_Cap ~ Corporate.equities.and.mutual.fund.shares + Labor.share +
##     Net.worth + Output + (1 | Category)
##  observations: 120
## ------
##                                           Median MAD_SD
## (Intercept)                               -80.6   20.2
## Corporate.equities.and.mutual.fund.shares   0.0    0.0
## Labor.share55.3                            31.6    7.2
## Labor.share56.0                            24.6    6.3
## Labor.share56.3                            17.8    7.8
## Labor.share56.5                             7.3    8.5
## Labor.share56.6                            14.2    7.6
## Labor.share56.9                            29.7    8.1
## Labor.share57.1                            12.6    7.6
## Labor.share57.2                            10.3    7.2
## Labor.share58.8                            40.0    7.5
## Labor.share58.9                            37.8    7.3
## Labor.share59.3                            45.8    7.2
## Labor.share59.9                            18.0    6.7
## Labor.share60.4                            54.2    7.8
## Labor.share60.5                            37.7   10.5
## Labor.share60.6                            55.1   10.2
## Labor.share61.0                            77.0    9.4
## Labor.share61.3                            30.3    7.8
## Labor.share61.4                            49.2   10.8
## Labor.share61.8                            45.2   11.1
## Labor.share61.9                            88.4    7.8
## Labor.share62.0                            42.1   11.4
## Labor.share62.5                            56.2    9.2
## Labor.share62.8                            78.8    8.2
## Labor.share62.9                            34.2   11.4
## Net.worth                                   0.0    0.0
## Output                                      1.8    0.2
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 10.1    0.8
##
## Error terms:
##  Groups   Name        Std.Dev.
##  Category (Intercept)  3.1
##  Residual             10.1
## Num. levels: Category 4
```

```
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Conclusion Both models were unsuccessful in determining whether or not the four predictors influence the market cap value and growth. Although both models gave interesting results, the coefficients were not appropriate and the residuals were not acceptable. It is possible a simple linear regression model would work best with this kind of data. Although the EDA showed trends in equity and wealth of the top 1% of income earners and the increase in market cap over the years, the propensity score modeling and linear regression method as well as the multilevel linear regression model did not show a causation between the predictors and the outcome. They were both inconclusive in determining whether or not there is causation between the wealth and equity of the 1% income earners and the increase in the market cap.

Appendix

In this appendix you will find additional plots of the residuals and qqplots for the multilevel linear regression model as well as data sources and explanation of units.

Data Sources:

- US Labor Market
  - U.S. Bureau of Labor Statistics: https://www.bls.gov/lpc/data.htm
- US Market Cap as Percent GDP
  - World Bank Database: https://data.worldbank.org/indicator/CM.MKT.LCAP.GD.ZS?end=2018&locations=US&start=1980&view=chart
- Wealth Data on the Top 1%
  - The Federal Researve: https://www.federalreserve.gov/releases/z1/dataviz/dfa/distribute/chart/#quarter:0;series:Net%20worth;demographic:networth;population:1,3,5,7;units:levels;range:2005.2,2020.2

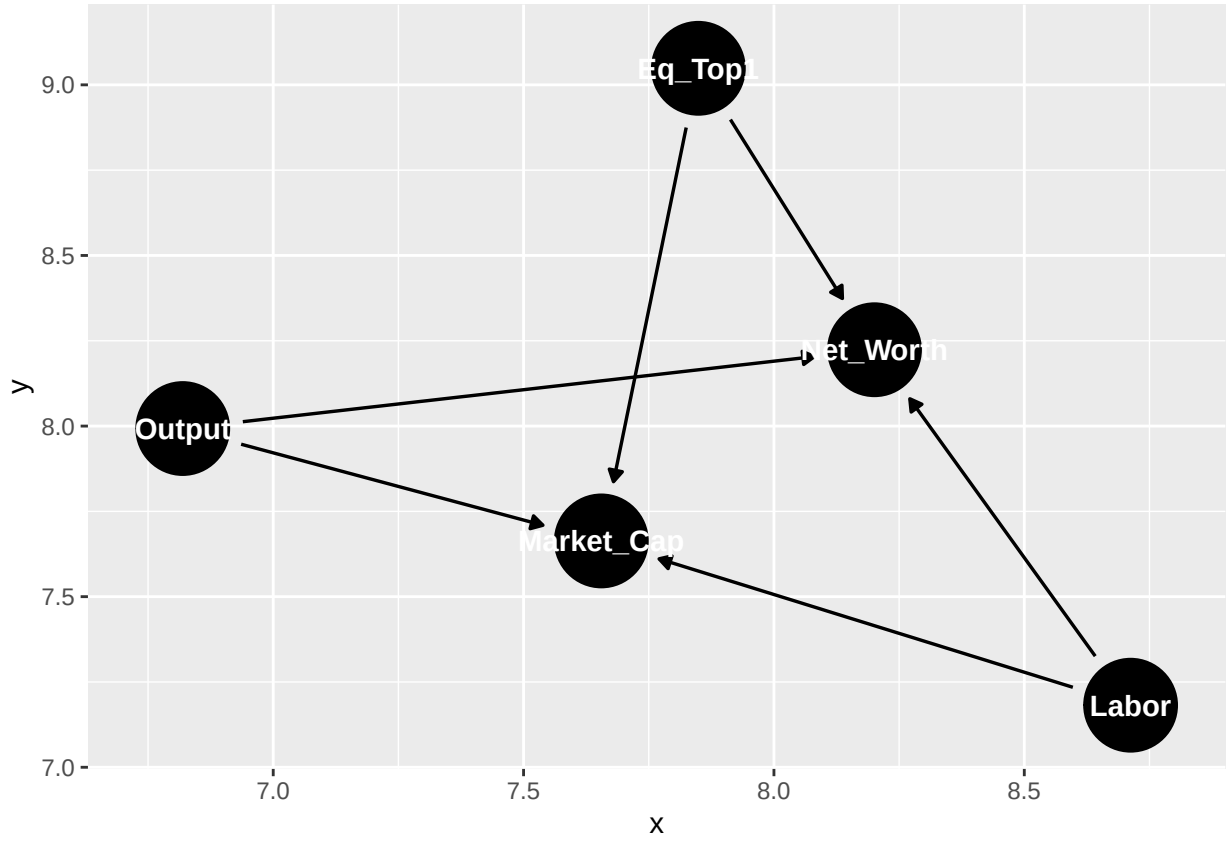
Calculations and Units Both market capitalization and labor share are expressed in percentages and on a per year basis:

- Labor Share = (compensation earned in dollars / total current dollar output produced) * 100
- Market Capitalization as % of GDP = (US Market Cap. in US dollars / GDP in US dollars) * 100
- Net Worth: the total value of assets a person owns minus the total liabilities he or she owes.

Diagram of Assumptions for Causal Relationship

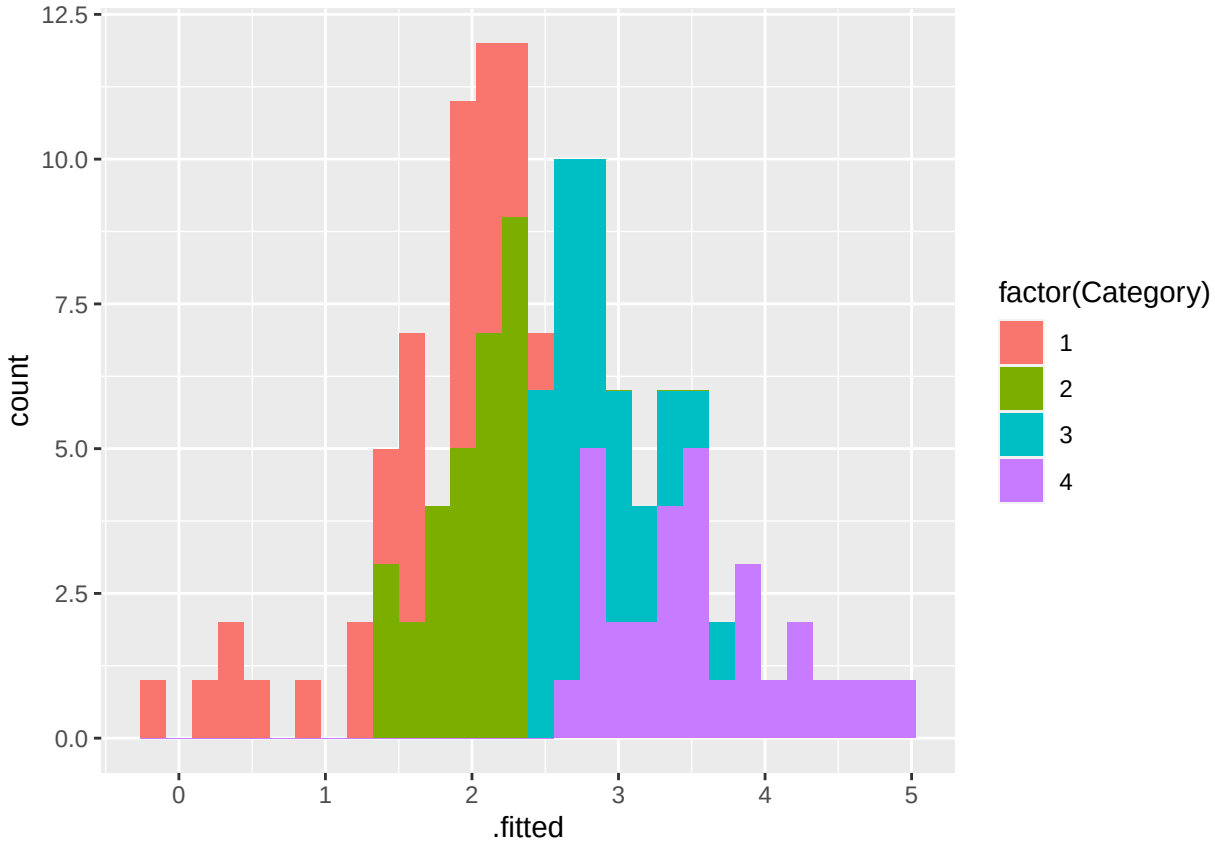
```
##
## Attaching package: 'ggdag'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```



Histogram of the Propensity score

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


12.5
10.0
7.5
5.0
2.5
0.0
count
0
1
2
3
4
5
.fitted
factor(Category)
1
2
3
4

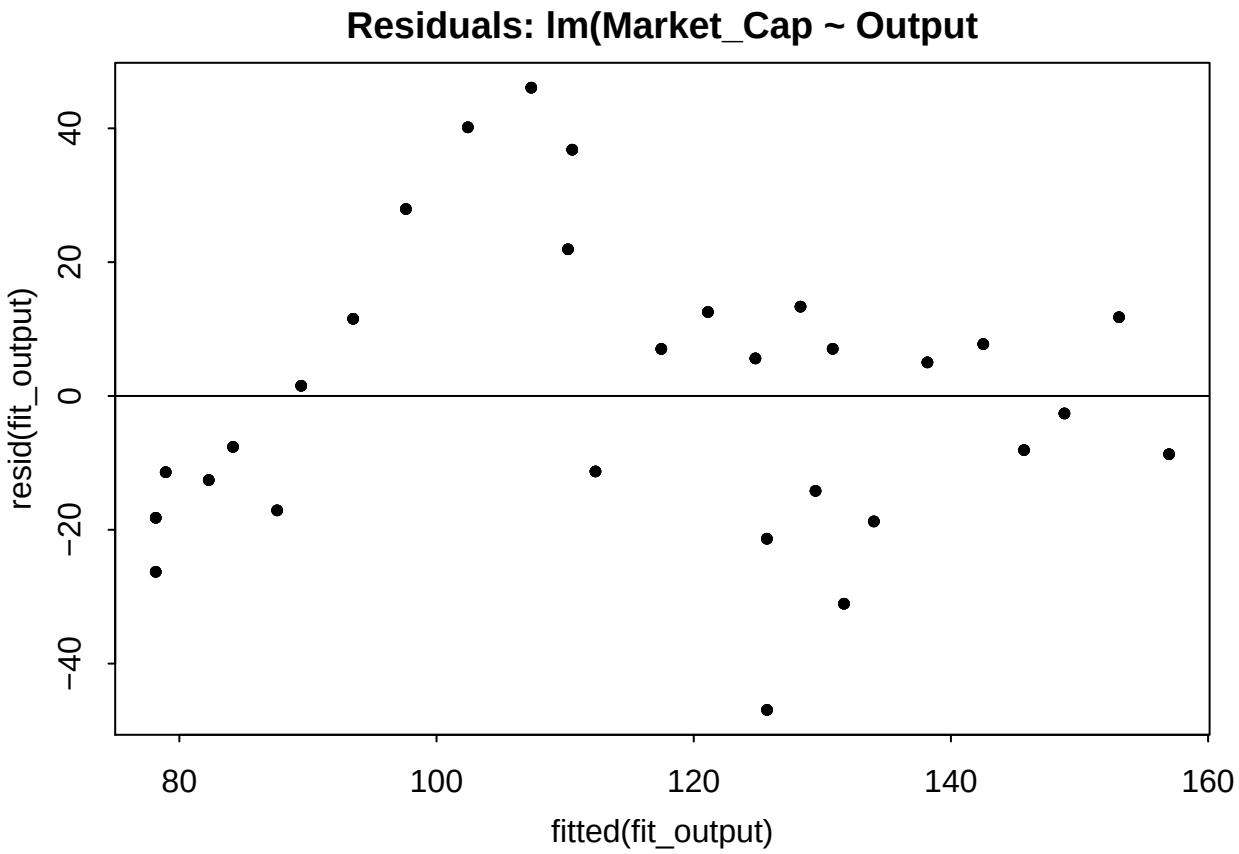
Residuals for the Multilevel model

Below are residuals of other simple linear regression models:

Residuals for Output:lm(Market_Cap ~ Output")

This is a residual plot for a linear regression model I tried with the data. There does seem to be a pattern. Output may not be linearly related to market cap.

```
##
## Call:
## lm(formula = Market_Cap ~ Output, data = data_frame2, refresh = 0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.923 -14.189  -0.544  11.769  46.077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.39676    8.53285   1.687   0.0942 .
## Output       1.19225    0.09793  12.175   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.22 on 118 degrees of freedom
## Multiple R-squared:  0.5568, Adjusted R-squared:  0.553
## F-statistic: 148.2 on 1 and 118 DF,  p-value: < 2.2e-16
```



Residuals for Labor Share: lm(Market_Cap ~ Labor_share")

```
##
## Call:
## lm(formula = Market_Cap ~ Labor.share, data = data_frame2, refresh = 0)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -32.3    0.0    0.0    0.0   32.3
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     100.6313     6.8462  14.699  < 2e-16 ***
## Labor.share55.3  42.5624     9.6819   4.396 2.87e-05 ***
## Labor.share56.0  32.1476     8.3848   3.834 0.000227 ***
## Labor.share56.3  45.5809     9.6819   4.708 8.52e-06 ***
## Labor.share56.5  47.6284     9.6819   4.919 3.64e-06 ***
## Labor.share56.6  36.9565     9.6819   3.817 0.000240 ***
## Labor.share56.9  64.2137     9.6819   6.632 2.01e-09 ***
## Labor.share57.1   3.7174     9.6819   0.384 0.701868
## Labor.share57.2  14.6266     9.6819   1.511 0.134182
## Labor.share58.8  29.7770     9.6819   3.076 0.002744 **
## Labor.share58.9  37.2214     9.6819   3.844 0.000218 ***
## Labor.share59.3  41.0228     9.6819   4.237 5.24e-05 ***
## Labor.share59.9   5.5823     8.3848   0.666 0.507177
```

```
## Labor.share60.4  14.1450     8.3848   1.687 0.094889 .
## Labor.share60.5 -30.1281     9.6819  -3.112 0.002456 **
## Labor.share60.6  -9.6332     9.6819  -0.995 0.322278
## Labor.share61.0  24.9306     9.6819   2.575 0.011569 *
## Labor.share61.3   0.4477     9.6819   0.046 0.963213
## Labor.share61.4 -24.0689     9.6819  -2.486 0.014667 *
## Labor.share61.8 -30.9133     9.6819  -3.193 0.001911 **
## Labor.share61.9  47.3837     8.3848   5.651 1.66e-07 ***
## Labor.share62.0 -40.6795     9.6819  -4.202 5.98e-05 ***
## Labor.share62.5  -0.7841     8.3848  -0.094 0.925689
## Labor.share62.8  46.7276     9.6819   4.826 5.30e-06 ***
## Labor.share62.9 -48.7552     9.6819  -5.036 2.26e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.69 on 95 degrees of freedom
## Multiple R-squared:  0.8515, Adjusted R-squared:  0.814
## F-statistic: 22.69 on 24 and 95 DF,  p-value: < 2.2e-16
```