How Do We Decide How Much to Save?
Consumer Preferences on Savings Rate in the US

Cem Tener
STAT 216
12/13/2021

1. **Introduction:**

The GDP of an economy is determined by three factors: capital, labor, and technology. However, there is much more that affects economic growth. Savings is the long-term "fuel" of an economy. Savings rate is very crucial to find the optimum measure of spending, investing, and spending to maximize utility . Also, consumers should not spend all their money because then they would not have any funds to invest and grow the economy.

Central banks do not have a direct control over savings rate because it all depends upon consumer preferences. However, they do have tools that can indirectly affect consumer preferences such as interest rates, inflation, expected inflation, and employment. For this reason, I was really interested in researching this. I believe it is possible to build up a model that predicts consumer preferences for savings according to certain economic factors.

This paper will first examine the relationship between the given variables and savings rate between 01/2004 to 12/2019. Then, a parametric OLS regression and nonparametric rank-based regression to build multiple regression models for saving rate will be presented. Finally, I will explore how saving preferences were affected during the pandemic by an umbrella alternative and a quadratic regression model.

1.2 Variables:

FRED (Federal Reserve Economic Data) contains all the economic measures for the US economy. I selected five variables that are most relevant and significant in macroeconomics for the model to predict savings rate. I choose inflation, expected inflation, 5 year Treasury interest rate, income, and GDP.

> Inflation: I used the Personal Consumption Expenditure Price Index (PCEPI) as a measure of inflation because that is the index FED sets their inflation target up.

> Expected Inflation: 5-Year, 5-Year Forward Inflation Expectation Rate is a measure of what participants in an economy thinks inflation will average in 5 years. It is calculated by measuring the difference of yield on bonds for different maturities.

> Interest Rates: Higher interest rates would attract deposits and investments. I choose a 5 year Treasury bond as a measure of interest rate.

> Income: Income is the core of savings because consumers choose how much of their income they want to save. I used income per capita because we would like to measure the change in income per person to see the effect on the savings rate of each consumer.

> GDP: is a measure of how big an economy is. It accounts for consumption, investment, government spending, and net exports (exports minus imports)

*Data for regression modeling is monthly from 01/2004 to 12/2019.

**Data for umbrella alternatives for pandemic is quarterly from Q4 of 2019 to Q4 of 2020.

2. **Correlation Testing**

My interest in this project is to analyze consumer preferences of savings in the US. Correlation tests show the statistical relationship between two variables involved in the bivariate structure. The goal is to test the independence of two variables and, if it exists, assessing the type and degree of the dependency. We have n bivariate observations, meaning there are n pairs of (X_k, Y_k) in our data. For nonparametric tests, our only assumption is that the (X,Y) pairs are mutually independent and identically distributed according to some continuous bivariate population. For the parametric test, the X and Y distributions should be continuous and normal. I will discuss and contrast the parametric Pearson correlation test with non-parametric Kendall and Spearman correlation tests.

2.1 Correlation of Major Economic Variables with Savings Rate in the US

The table below shows the test stat, sample estimate and p-value of each Pearson, Spearman, and Kendall procedure for the correlation between the savings rate and our variables.

	Savings vs Inflation	**Savings vs income**	**Savings vs interest rates**	**Savings vs gdp per capita**	**savings vs expected inflation**
Pearson					
Test stat	t = -2.922, df = 62	t = 9.3588, df = 62	t = -8.2525, df = 62	t = 7.8009, df = 62	t = -2.4979, df = 62
p-value	0.004847	1.811e-13	1.445e-11	8.77e-11	0.01516
sample estimate (cor)	-0.347912	0.7651957	-0.7235033	0.7038004	-0.3023785
Spearman					
Test stat	S = 61140	S = 6867.5	S = 70587	S = 8870.2	S = 56734
p-value	0.001066	< 2.2e-16	6.024e-08	3.361e-15	0.01645
sample estimate (rho)	-0.399734	0.8427767	-0.6159995	0.7969266	-0.2988503
Kendall					

Test stat	z = -2.846	z = 7.4483	z = -4.6197	z = 6.8919	z = -1.9186
p-value	0.004427	9.452e-14	3.843e-06	5.505e-12	0.05503
Estimate (tau)	-0.244831	0.6407468	-0.3974126	0.5928777	-0.1650484

2.2 Correlation Results and Comparison

In calculating the correlation coefficient Perason and Spearman have a similar way they correlate normally distributed data. On the other hand, Kendall is a test of strength of dependence which means that one variable can be written as a linear function of the other variable.

Looking at the table, we can observe that inflation, income per capita, interest rates and GDP per capita are significantly correlated with savings rate. However, for expected inflation, there is a conflict between the conclusions of Kendall test, and Pearson and Spearman. It is not surprising that Pearson and Spearman methods have similar conclusions since their calculations are parallel. At a 95% significance level, Kendall test does not conclude a significant correlation between expected inflation and savings rate. Even though the p-value is extremely close to our significance level, we cannot reject the null hypothesis. Kendall's test is most powerful when we have a small sample or outliers in the data. Since we have a large sample size and no outliers, and the p-value is on the edge, it is possible to rely on Spearman's and Pearson's conclusions.

3. Regression Modeling

The main goal of this project is to understand the saving preferences according to major economic measures. In order to achieve this goal, constructing a multiple regression model is the most suitable. A multiple regression model has the following form:

$Yi = \xi + \beta 1x1i + \beta 2x2i + \ldots + \beta pxpi + ei,\ i = 1, \ldots, n$

where,

ξ := intercept parameter
~xi = (x1i, x2i, . . . , xpi) := vectors of known constants
βk := slope parameter for kth predictor variable
ei:= residual of the ith observation.

In this research, I used parametric Ordinary Least Squared (OLS) and the nonparametric regression method based on the Jaeckel-Hettmansperger-McKean test statistic. In both models, I used the backward elimination method where I started with including all the variables in the model. Then, I eliminated the least significant variables (which have the highest p-value) from the model until I had the most significant model with the least variables.

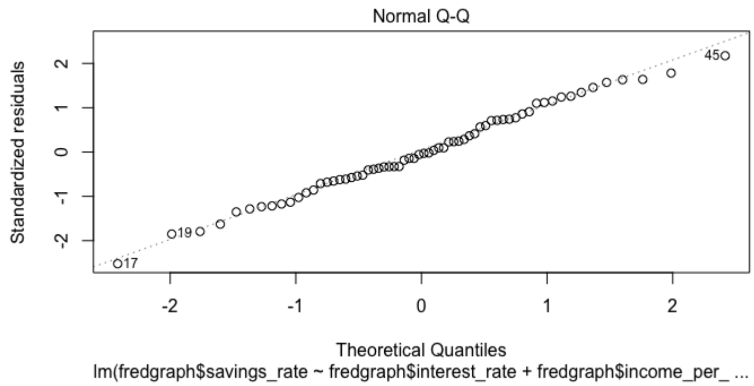
3.1 Ordinary Least Squared Parametric Model for Savings Rate in the US

This parametric method has conditions to run the model. The residuals should have constant variance, follow a normal distribution, and be independent from each other.

3.1.1 Checking the Conditions

Linearity-Constant Variance:
By looking at the Residuals vs Fitted plot, it seems that the residuals are distributed nicely around 0. However, residuals do not follow a constant variance and fail to satisfy linearity conditions.

Normality:
Second, we look at the normal normal quantile (QQ) plot to check for normality. Even though we have a couple outliers, we can confirm that our data is distributed normally.



I got the following model for savings rate by backward elimination method.

~~~~~~

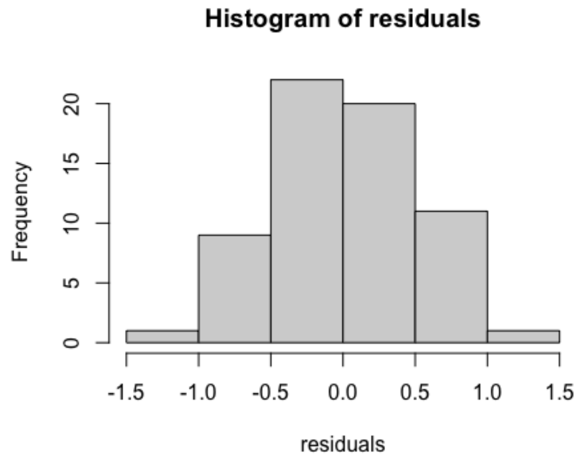
```
Call:
lm(formula = fredgraph$savings_rate ~ fredgraph$interest_rate + 
    fredgraph$income_per_capita + fredgraph$gdp_per_capita)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.15364 -0.28798 -0.01909  0.33424  1.00715 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  3.9554989  0.6114202   6.469    2e-08 ***
fredgraph$interest_rate     -0.2953492  0.0782496  -3.774  0.00037 ***
fredgraph$income_per_capita  0.0017727  0.0001318  13.446  < 2e-16 ***
fredgraph$gdp_per_capita    -0.0012860  0.0001057 -12.171  < 2e-16 ***
```

Finally, even though the conditions are not met perfectly, I built up the OLS model. I got a significant model that can make accurate estimates. According to my model interest rates, income per capita, and GDP per capita are the most significant variables to predict the savings rate in the US. Surprisingly, interest rates and GDP growth show a negative impact on savings in our model. However, with small p-values, they contribute to our model significantly. Also, it follows the economic sense that as income increases, consumers tend to save more. The Multiple R-squared associated with the model is 0.931, meaning that our model can explain 93.1% of variability can be explained by the model's variables.
~~~~~~

3.2 Non-parametric Model for Savings Rate in the US



Looking at the histogram, we can see that residuals are distributed symmetrically at 0 which satisfies the condition for the non-parametric procedure.
I got the following model for savings rate by backward elimination method.

```
Call:
rfit.default(formula = fredgraph$savings_rate ~ fredgraph$interest_rate +
    fredgraph$income_per_capita + fredgraph$gdp_per_capita)

Coefficients:
                               Estimate  Std. Error  t.value   p.value
(Intercept)                  4.12030556  0.65079300   6.3312 3.415e-08 ***
fredgraph$interest_rate     -0.27819503  0.08306733  -3.3490  0.001406 **
fredgraph$income_per_capita  0.00179318  0.00013996  12.8121 < 2.2e-16 ***
fredgraph$gdp_per_capita    -0.00130522  0.00011216 -11.6368 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Notice that we have the same variables which are significant in our model in predicting savings rate in the US. The signs of the estimates are identical with the OLS model and have similar values. Multiple R-squared: 0.8400825 which means that we can explain approximately 84% of the variance in our model with the model's variables. Even though both of the models can explain the majority of the variance in the data, the non-parametric model is weaker in terms of estimating savings rate. However, we are more confident that the conditions for non-parametric procedures are met.

3.3 Comparison between Parametric and Non-Parametric Models

I will use root-mean-square error statistics to measure the distance between the observed and true values. The model that has a smaller RMSE value is more accurate in predicting the true value according to the model's variables. The smaller RMSE a model has, its accuracy increases in predicting the dependent variable. The formula for the statistic is:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{n}}.$$

Parametric OLS Model

Savings Rate = 3.955499 -0.295349 * interest rate + 0.001773 * income per capita -0.001286 * gdp per capita

Rank-Based Non-Parametric Model

Savings Rate = 4.120305557 -0.278195032 * interest rate + 0.001793176 * income per capita -0.001305218 * gdp per capita

DATE	True Savings Rate	OLS Prediction	Non-parametric Prediction	DATE	True Savings Rate	OLS Prediction	Non-parametric Prediction
1/1/04	4.833	4.516	4.517	1/1/12	7.933	8.007	7.962
4/1/04	5.267	4.466	4.472	4/1/12	8.567	8.243	8.198
7/1/04	4.800	4.298	4.298	7/1/12	7.833	7.574	7.516
10/1/04	4.900	4.555	4.550	10/1/12	9.933	9.553	9.514
1/1/05	3.067	3.083	3.064	1/1/13	5.800	6.101	6.020
4/1/05	2.833	3.256	3.239	4/1/13	6.367	6.275	6.202
7/1/05	2.467	3.053	3.037	7/1/13	6.400	5.749	5.683
10/1/05	3.067	3.099	3.088	10/1/13	5.933	5.344	5.271
1/1/06	3.900	3.344	3.333	1/1/14	6.667	6.486	6.429
4/1/06	3.700	3.320	3.314	4/1/14	7.033	6.536	6.473
7/1/06	3.200	3.356	3.350	7/1/14	7.133	6.392	6.327
10/1/06	3.433	3.276	3.267	10/1/14	7.367	6.792	6.737
1/1/07	3.567	3.558	3.548	1/1/15	7.933	6.936	6.879
4/1/07	3.633	3.536	3.525	4/1/15	7.500	6.754	6.688

7/1/07	3.200	3.291	3.276	7/1/15	7.333	6.811	6.752
10/1/07	3.000	3.622	3.592	10/1/15	7.400	7.071	7.017
1/1/08	3.400	4.562	4.523	1/1/16	7.500	7.378	7.323
4/1/08	5.100	5.672	5.648	4/1/16	6.900	7.097	7.028
7/1/08	3.933	4.798	4.772	7/1/16	6.800	7.097	7.027
10/1/08	5.900	5.233	5.248	10/1/16	6.800	7.059	6.990
1/1/09	5.900	5.797	5.795	1/1/17	7.100	7.506	7.442
4/1/09	6.900	6.658	6.662	4/1/17	7.500	7.837	7.776
7/1/09	5.200	6.021	6.017	7/1/17	7.500	7.663	7.598
10/1/09	5.467	5.947	5.929	10/1/17	7.000	7.274	7.204
1/1/10	5.700	6.229	6.210	1/1/18	7.533	7.697	7.633
4/1/10	6.467	6.488	6.470	4/1/18	7.500	7.469	7.400
7/1/10	6.467	6.431	6.404	7/1/18	7.600	7.758	7.693
10/1/10	6.300	6.476	6.442	10/1/18	7.867	8.049	7.991
1/1/11	6.900	7.438	7.418	1/1/19	8.567	8.263	8.198
4/1/11	6.600	7.173	7.139	4/1/19	7.433	7.509	7.426
7/1/11	6.767	7.514	7.477	7/1/19	7.167	7.470	7.379

3.4 Regression Conclusion and Comparison

Looking at the table, both parametric and non-parametric regression models look successful in predicting the savings rate in the US between 01/2004 and 12/2019. To compare the models, the OLS model has a 93.1% R^2 value which is higher than the non-parametric models that has approximately 84% R^2 value. However, my parametric model has a RMSE value of .37 and the non-parametric model has .3639. The non-parametric model has a lower RMSE than the OLS model. Even though the RMSE values are very close to each other, the non-parametric model is slightly more accurate in predicting savings rates in the US. Even though the OLS model has a higher R^2, it makes sense that OLS is not as accurate as the non-parametric model because its conditions are not met perfectly.

4. Effect of Pandemic on Savings Rate in the US

With the shutdown of the whole world during the 2020 pandemic, our daily habits evolved accordingly to the COVID protections. We have many new procedures at school, work, traveling etc. As a follow up of this research, I would like to analyze the effect of the pandemic on savings rate in 2020.

During recessions, savings rates tend to go up because of the scarcity in consumers. Furthermore, during the lockdown, there weren't a lot of goods and services available. Therefore, I expect to have an increase in the savings rate and then a gradual drop back to normal level as pandemics effects diminish. I will use a Umbrella Alternative-Peak Unknown test to see if there are differences between the months of 2020. I am using the Peak Unknown method because it is not certain to know when consumers started to prefer savings over spending. It can be right in the beginning of the pandemic or a month after lockdown when they really cannot spend money. I will also construct a quadratic regression model to predict the preference of consumers and see what kind of trend the regression model gives me.

4.1 Umbrella Alternative Test Results

For the sample from December 2019 to December 2020, I had the following hypothesis;

H_o: [all the treatments are equal] vs H_1: [T_1<=...<=T_{p-1}>=...>=T_k, with at least one strict inequality, for some p{1,2,...,k}]

Conclusion:

With Mack-Wolfe Peak Unknown A*(p-hat) Statistic: 3.6836> critical value for Mack-Wolfe Peak Unknown A*(p-hat)=2.3112508176 (at a significance level of 95%), we have sufficient evidence to reject the H_o and conclude that the savings preferences of the consumers followed an umbrella trend.

4.2 Quadratic Regression Model

With the same method, backward elimination, I got the following model.

~~~~~~

```
Call:
lm(formula = pandemic_new$PSAVERT ~ pandemic_new$PCE_PC1 +
pandemic_new$A229RC0_PC1 +
    pandemic_new$DFII5 + income_per_capita3)

Residuals:
      Min        1Q    Median        3Q       Max
-0.151376 -0.042516 -0.007971  0.049667  0.093771

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               9.174741   0.077509 118.370 2.90e-14 ***
pandemic_new$PCE_PC1     -0.797252   0.009703 -82.169 5.37e-13 ***
pandemic_new$A229RC0_PC1  0.821325   0.028307  29.015 2.15e-09 ***
pandemic_new$DFII5        0.283049   0.062027   4.563  0.00184 **
income_per_capita3       -0.006824   0.001845  -3.699  0.00605 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```
~~~~~~

Model:

Savings Rate in the US between 12/2019 and 12/2020 = 9.174741- 0.797252 * inflation + 0.821325 * income_per_capita + 0.283049 * interest_rates -0.006824 * income_per_capita^2

In my quadratic model for savings rate during the pandemic, inflation, income_per_GDP, interest rates, and square of income_per_GDP were the significant variables. Surprisingly, the model has an R^2 value of 99.99%, which means that my quadratic regression model can approximately explain all the variability in the data.

DATE	**True Savings Rate**	**OLS Prediction**
12/1/19	7.3	7.208
1/1/20	7.8	7.951
2/1/20	8.3	8.206
3/1/20	13.1	13.143
4/1/20	33.8	33.817
5/1/20	24.8	24.750
6/1/20	19.3	19.312
7/1/20	18.7	18.674
8/1/20	15.0	15.008
9/1/20	14.3	14.366
10/1/20	13.6	13.652
11/1/20	13.0	12.970
12/1/20	14.0	13.943

Looking at the table, we can see that the quadratic regression model is highly successful in predicting the true savings rate during the pandemic. The RMSE statistic of the model is .0536 which is extremely small and shows the accuracy of the model.

4.3 Comparison of the Umbrella Alternative and Quadratic OLS Model

Both the Umbrella Alternative and the Quadratic OLS model were successful in testing the effect of the pandemic on savings rate preference. The Umbrella Alternative can only test the treatment effects but cannot provide an estimate for the Q4 of 2019 to Q4 of 2020. It can only show the trend in the data. On the other hand, the OLS model is very accurate in predicting the savings rate during the pandemic. Visually, we can see that the data follows an umbrella model. However, it does not provide any statistical value whether the savings was affected by the pandemic. Further, even though the quadratic model I built is extremely successful, it can only predict for such a small data. These two tests are not identical and do not serve the same hypothesis. We would need to build a new model to predict the estimates for new data which would again not have statistical evidence about the treatment effect. It would be best to use the

Umbrella Alternative to test treatment effects and to estimate savings rate given the variables, a quadratic model would be the best. This way we would have statistical evidence for treatment effect and estimates for savings rate.

5. Conclusion:

Overall, I underlined how consumers decide on how much of their income they are going to save in the US according to given macroeconomic variables. Correlation values for income per capita, GDP per capita, inflation, expected inflation, and interest rate show that each of these macroeconomic variables are correlated with consumer preferences in the US. Only Kendall's test did not have a parallel conclusion with the Spearman and Pearson correlation test. However, this can be disregarded since we do not have a small sample and outliers. Both parametric and non-parametric tests gave parallel conclusions.

It was interesting that the OLS model I built had a higher R^2 and explained more of the variability in the data than the non-parametric model, but it was less accurate in predicting the true value of savings rate. By nature, OLS tries to get the line with the least squared distances to the data. Therefore, I was expecting the OLS to be more accurate. However, since the conditions were not fully met, the non-parametric model was more accurate in predicting by looking at the RMSE values.

With the statistical evidence from the Distribution Free Umbrella Alternative-Peak Unknown, we rejected the H_o and concluded that there was an umbrella design in the savings rate in the US during the pandemic. We were able to visualize this with the quadratic OLS model and estimate the savings rate given the significant variables. However, the quadratic model did not give any statistical evidence on treatment effects.

This research opens the doors for further research. It would be very interesting and useful to see what factors might push consumers to save at the "golden ratio" which maximizes steady state level of the growth of consumption as in the Solow Model. This can give knowledge to economists in which macroeconomic levels consumers will save at the golden ratio to maximize the steady state of the growth of consumption.

Works Cited

Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods*. John Wiley and Sons Ltd.

Federal Reserve Bank of St. Louis. (n.d.). *Federal Reserve Economic Data: Fred: St. louis fed*. FRED. Retrieved December 15, 2021, from https://fred.stlouisfed.org/