Evaluating Household Debt

George I. Kaveladze, Jason I. Chavez, John H. Tang, Khwala A. Abdulgader, Oluwagbemiga M. Adeosun
{gkaveladze3,jchavez8,jtang338,kabdulgader3,oadeosun7}@gatech.edu

ABSTRACT

The ability to accurately predict economic expansion or contraction is shown to be heavily reliant on household debt. The correlation between economic downturns and high household debt levels is high, with household debt to GDP ratios having slightly higher predictive power. Our household debt dashboard will provide users with the ability to observe the effects that changes in economic factors has on household debt and economic health, by proxy. This approach will differentiate itself by predicting a factor that plays a large role in economic health, rather than attempting to predict economic health as a whole, allowing the model to account for more nuanced prediction factors.

KEYWORDS

Datasets, Household Debt, Economic Visualization, Household Debt to GDP Ratio, GDP Ratio, Interest Rate

1 INTRODUCTION

The field of macroeconomics has produced a wide variety of research into the predictability of economic expansion and contraction. Interest in this research field has increased over the years with many recent recessions providing a swath of data that can be analyzed and evaluated to determine the level of predictive power. The Great Recession of 2008, for instance, highlighted a set of factors that could be used to predict economic fluctuations, including recessions and market corrections. Household debt, for example, was shown to have a negative impact on the GDP despite providing a short-term stimulus[7]. Accurately predicting this value could provide insights into of economic fluctuations and financial instability associated with government policies and consumer spending. Research conducted by Zabai[15] has indicated that there is a direct correlation between economic health and household debt. Our goal for this project is to provide an interactive dashboard that allows users to observe correlations between household debt and other economic factors, along with predicting household debt given user-determined values for factors like consumer price index and yields on US Treasury 10 year bonds.

Since a large and rapid increase in household debt points to economic slowdown[11], this information can help users understand the impact their spending habits have. Similarly, the ability to adjust values and see outcomes could help legislators better understand the impact their policies have on the economy[6].

2 PROBLEM DEFINITION

When a decline in the economic activity has been observed for multiple consecutive months a recession is to be expected in the United States. This directly destabilizes households, small businesses, and large corporations. US policy makers will utilize stimulus packages to mitigate long-term consequences but a spike in the unemployment rate and defaults on mortgages still occurs in the near term. Recessions have been observed to follow a business cycle. The ability to predict when the next economic downturn will occur allows corporations and US policy makers to properly plan and prevent a greater negative impact to the economy. As previous research has indicated there is a strong correlation between economic downturns and a high household debt[11], our dashboard will allow users to observe the effects that changes in economic factors have on household debt and economic health, by proxy. Users will have the opportunity to adjust certain factors and observe the consequences of these values through the output of our Machine Learning algorithm that displays the predicted household debt.

3 SURVEY

Our dashboard will leverage several studies performed on the correlative effects of household debt and economic downturns.

Studying The Great Recession
Many studies have been conducted to better understand the factors that lead to the 2008 recession, including Chakrabarti et al.'s evaluation of household debt and savings during the recession[3], Nyman et al.'s Machine Learning approach to understanding the great recession[14], and Mian et al.'s observation of household leverage and the recession[12]. All articles show the high correlation between economic downturns and increased household debt. We've used these articles to back our belief in the utility of household debt to predict recessions.

Macroeconomic effects of Household Debt
General evaluations of the macroeconomic effects have been outlined in great detail in papers like Alter et al.'s global perspective on household debt effects[1], Friedman's theory of the consumption function[5], Kim's empirical analysis of the effects of household debt[7], Lombardi et al.'s evaluation of the real effects of household debt[9], Mian et al.'s observations on household debt and worldwide business cycles[13][11], and Filardo's assessment on the reliability of prediction models[4]. Each of these articles provide a backing for the global reach our proposed predictions can have, and provide a solid background on which correlations should receive particular attention. We'll expand this information by including more concentrated variables that relate specifically to household debt.

Policy Impacts on Economies and Household Debt
If the provided proof of the relation between household debt and economic health have provided justification for our dashboard, studies on legislative effects provide motivation for creating the dashboard. Garber et al.'s study of Brazil's 2014 recession[6] and Guggenheim Investments' look into the effects rate cuts will continue to have in the US economic health[2] provide a basis for which variables our dashboard will allow users to adjust.

Current State of Debt
A major push to produce our dashboard has come from research that reveal household debt is steadily climbing. From Li's evaluation on the economics of student loans[8], to Mian et al.'s study on the household leverage crisis[10] and Zabai's assessment on recent household debt developments[15], we've come to realize how much household debt continues to grow. Leveraging this information, we hope to reveal what factors might be leading to this troubling trend.

4 PROPOSED METHOD

The following sections outline the methodology we have taken to produce our application:.

4.1 Dashboard

The Household Debt dashboard will provide a multifaceted approach to understanding the effects that various factors have on how citizen's borrowing practices. This will be accomplished by, first, allowing users to observe visualizations that convey the relations between a subset of factors and household debt. A household debt increases have been shown to be directly correlated with economic health[13], these visualizations will allow users to better understand the contributions that this debt has on a macroeconomic level and help curb the significant growth in household debt[1].
Additionally, users will have the ability to adjust the values associated with these highly correlative factors and observe the resulting household debt predictions. Our predictions will be presented both textually and graphically as we explore visualizations that demonstrate the prediction's relation to GDP.

4.2 Predictions

In order to provide significant predictions, we will gather data from a variety of sources and evaluate variable importance for predicting household debt. We have collected 29 economic data variables and completed a correlation analysis using R. While other models have been created using the random forest algorithm[14], we compared predictions from multiple machine learning models and found that linear regression produced the best overall predictions.

4.3 Technologies

The application's user interface is a responsive single page application (SPA) based on AngularJS and Bootstrap. The graphs were generated using D3.js for visualization. We wrote extensive Scala and Apache Spark code to clean, transform, and insert our data into a SQLite database. We used the same technology to then aggregate the data and export it to a csv file which was then used to train and test our Azure ML Studio algorithms. We containerized our solution using Docker to allow users to run our cleaned data for their own exploration. Our website will be hosted on Heroku and has a Node.js back-end. The web application makes REST API calls to the Azure ML Studio for predictions. Our web application is hosted at https://teamfed-project.herokuapp.com//landingPage.

We have designed our application so that it can be executed a local PC or on the cloud. The options to run our application are listed below.

Development server (Local Run without Docker)

- Start a command shell
- Navigate to the webapp directory.
- Run the command "npm install"
- Run the command "npm start"
- Start a browser and navigate to http://localhost:8080/#/landingPage

Docker server (Locally run latest code using Docker)

- Launch a command shell
- Navigate to the webapp directory
- Run the command "docker build -t teamfed ."
- Run the command "docker run -p 8080:8080 teamfed"
- Launch a browser and navigate to http://localhost:8080//landingPage

Deploy to Heroku (Prod Run)

- Install the HerokuCLI from https://devcenter.heroku.com/articles/heroku-cli
- Run heroku login to create a session to your heroku account.
- Now you can sign into Container Registry heroku container:login.
- Run heroku container:push web -a teamfed-project to create a docker image and push it to our Heroku repository.
- Run heroku container:release web -a teamfed-project to release the latest image and this will update the webapp.
- Run heroku logs –tail -a teamfed-project to see the logs and output from interacting with the webapp.
- Navigate to https://teamfed-project.herokuapp.com//landingPage and watch the as the logs change and acknowledge the interaction with the webapp.

4.4 Innovations

Our dashboard includes the following innovative approaches. Firstly, based on our extensive research, there is a limited amount of models focusing on predicting household debt only, even though it is a strong indicator to the economic health of the United States. Many financial institutions produce economic models using household debt as a primary variable, therefore we are going to focus on this specific variable rather than predict the larger economy. Secondly, we are trying to account for negative rates by using the German 10 year yield in our model. This will allow our model to include data that is typically not used in United States recession models as the data is unavailable, but provides a complete view to an economic downturn. Lastly, we are streamlining our dashboard by providing an API of our machine learning model to the front-end and having a Node.js back-end while containerizing our solution using Docker.

5 EXPERIMENTS AND EVALUATION

In order to validate our methodology when creating our household debt dashboard, we have compiled a list of steps that we have reviewed throughout our process. These steps are outlined as follows. Firstly, we have collected a large amount of data sets to determine which input variables will provide the highest predictive power as we believe this will allow us to have the best prediction performance. The data sets and their sources are listed in the Appendix under "Table 1 Data Sets". Secondly, we utilized multiple machine learning algorithms: Bayesian Linear Regression, Linear Regression, Boosted Decision Tree, Decision Forest Regression, and Neural Network Regression. This ensured that our study utilized the most suitable methodology when predicting household debt. Thirdly, we have included a user functionality feature as we believe allowing the user to enter various data elements will provide a greater understanding of the factors that impact household debt. Fourthly, we have been researching various economic dashboard visualizations to ensure we compartmentalize the information in an intuitive manner for the user. Lastly, we believe our front-end and back-end methodology is stream-lined as stated in the innovations section above.

We used R with the tidyverse, GGally, rpart, and caret libraries to help select the input variable with the highest correlation to household debt. We initially ran the correlation analysis against all of our input variables. See figure 5.1 below and "Table 2 Correlation For All Input Variables" in the appendix. This plot is the correlation matrix of all the variables we evaluated for this project. The major aspects of this plot that we evaluate are:

- The correlation each variable has with the response variable (Household Debt). We do this both visually and numerically, where the straighter the line visually, the higher the correlation value ought to be numerically
- Correlation of each predictor variable to one another. If a predictor variable is highly correlated with another variable, we needed to choose one of the two as to not over-weight the underlying data that both variables are explaining.

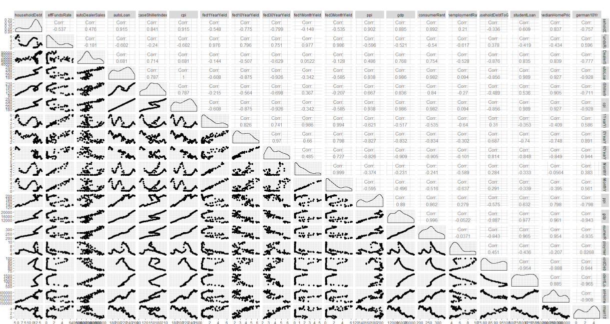

Figure 5.1 Correlation For All Input Variables

We dropped input variables which had less than 0.75 correlation to household debt or if the ML model showed that an interaction of variables was highly predictive. As shown in the Figure 5.2, both the Fed 3 Month Yield and Unemployment Rate had low correlations, but the interaction they had with other variables improved the overall model score. Using this graph as a guide, we reduced the number of input variables from eighteen to just nine. Our final list of input variables are the consumer price index, Treasury 3 month bond yield, Treasury 10 year bond yield, Treasury 30 year bond yield, consumer rent, German 10 year bond yield, SP Case Shiller Index, Producer Price Index, and the unemployment rate. Please see the appendix "Table 3. Correlation For Final Input Variables" for a larger correlation graph.

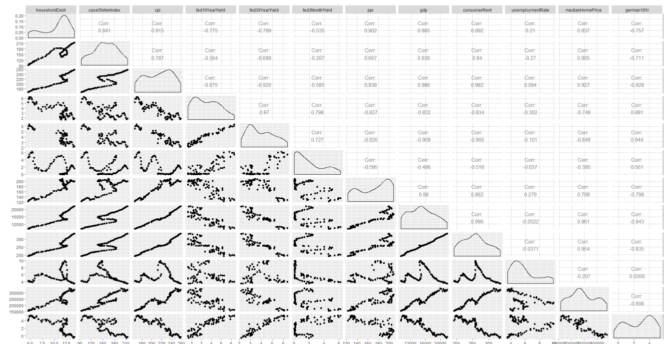

Figure 5.2 Correlation For Final Input Variables

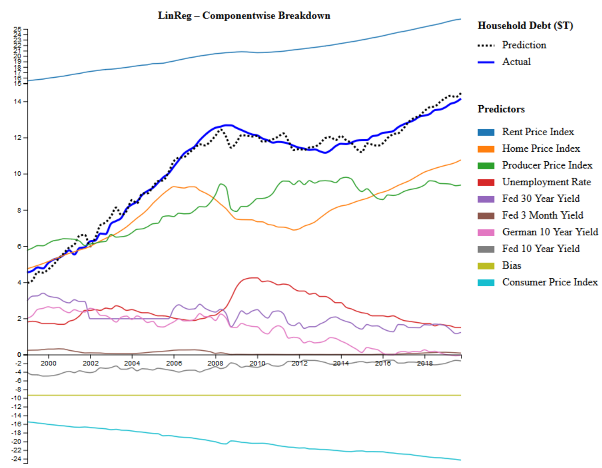

Figure 5.3 Prediction vs Final Input Variables and Actual Household Debt

We also had to determine that time scale our study should use. We had to decide if our model predicted household debt on a yearly, quarterly, or monthly basis. We were inspired by Nyman et al.'s Machine Learning approach to understanding the great recession[14] use of a quarterly time scale. Their Rain Forest model did a poor job on a daily and yearly scale but it worked quite well on a quarterly basis. We also created a violin plot to see how each quarter's household debt distribution stacked up against each other. This plot allowed us to see if a quarter time scale should be a variable used for prediction. As shown in the plot, each quarter's distribution closely matched that of the other quarters.

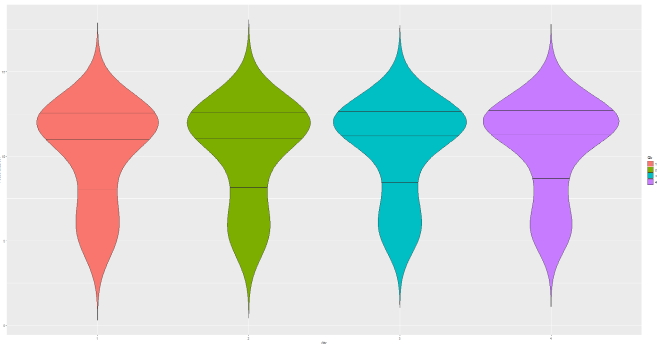

Figure 5.4 Violin Plot of Debt On Quarterly Time Scale

Another assessment we had to perform was the effectiveness of the number of input variables our study used. Did we used too many or too few? Did the study degrade as we added more input variables? To address these concerns we created a Lasso Regression plot of the mean squared error as input variables were added to the model. It was clear that adding variables did not degrade the performance of our model.

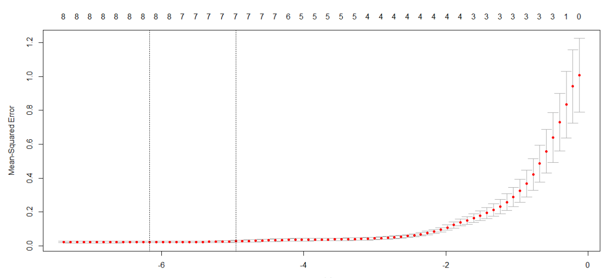

Figure 5.5 Lasso Regression Plot of Mean Squared Error vs Num of Input Variables

JASON TO-DO: Please explain Household Debt Auto Corr plot

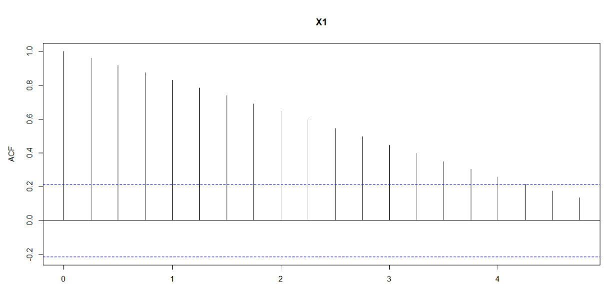

Figure 5.6 Household Debt Auto Correlation

We used a data split of 80 percent for training and 20 percent for testing. Our web application feeds the input variables to a Microsoft Azure Machine Learning algorithm using REST API calls. The calls are secured with a private key. We were shocked as we got back the predictions. We thought the more sophisticated machine learning algorithms such as Decision Forest Regression or Neural Network Regression would produce superior predictions. It turned out that the Linear Regression model created as good or better predictions. We committed to the Linear Regression model for our study.

Variable	Coefficient
Bias	-.937356
Fed 10 Yr Yield	-0.790002
Fed 30 Yr Yield	0.534679
German 10 Yr Yield	0.495695
Unemployment Rate	0.428106
Consumer Price Index	-0.0940645
Consumer Rent	0.0798328
Fed 3 Month Yield	0.0515536
Case Shiller Index	0.0503518
Producer Price Index	0.047038

Figure 5.7 Model Coefficients

Finally, we used the model to make a prediction for the four quarters of 2020.Please see Figure 5.8 for a plot of the Holt -Winters Exponential Smoothing model.

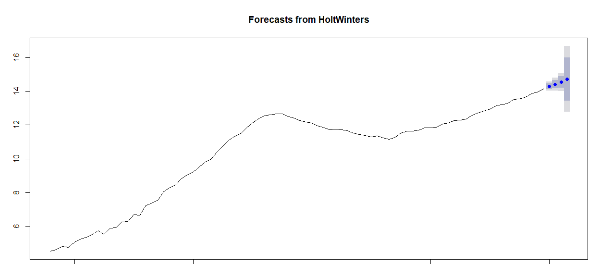

Figure 5.8 Holt-Winters Smooth Forecast

TO-DO: Need Final Web UI and write up

Please see the Appendix for information about the project's source code.

6 PLAN OF ACTIVITIES

We achieved success by following the plan below:

Activity	Completion Date (03/27)
Collect Data	03/07
Variable Exploration	03/20
Clean Data	03/21
Variable Selection	03/27
Progress Report	03/27
ML Algorithm Developed	04/03
Front-end Setup	04/03
Back-end Setup	03/31
Final Report	04/13

All group members have continued to contribute a similar amount of effort. The following activities were completed by the following teammates: Data Collection: John and Khwala; Variable Exploration: George; Data Cleansing: John and Jason; Variable Selection: Jason and Khwala; ML Algorithm: Jason; Front-end Set Up: Bemi and George; Back-end Set Up: Bemi; Progress Report: Khwala; Final Report: Khwala, Jason, and John.

7 CONCLUSIONS AND DISCUSSION

We intially started with 28 input variables. We knew that some of the variables were going to be eliminated during our variable selection phase. We were surprised that automobile loans, auto dealer sales, consumer rent, and student loans did not make the final list of input variables.

Graphing and analyzing exisisting data provided a clear near term trend of household debt. We were not suprised when our model predicted ever increasing household debt for 2020. We were very worried about what this meant for the econonmy. The model predicted a household debt value that was far greater than the value from the last recession of 2008. While we didn't know that the coronavirus would be the spark that would deflate the economy, our model clearly alerted us that the economy was going to tank.

REFERENCES

[1] Adrian Alter, Alan Xiaochen Feng, and Nico Valckx. 2018. Understanding the Macro-Financial Effects of Household Debt: A Global Perspective. (April 2018). https://www.imf.org/~/media/Files/Publications/WP/2018/wp1876.ashx

[2] Matt Bush, Scott Minerd, and Brian Smedley. 2019. Forecasting the Next Recession: Will Rate Cuts Be Enough? (September 2019).

[3] Rajashri Chakrabarti, Donghoon Lee, Wilbert Van der Klaauw, and Basit Zafar. 2015. Household Debt and Saving during the 2007 Recession. (April 2015). https://doi.org/10.3386/w16999

[4] Andrew J. Filardo. 1999. How Reliable Are Recession Prediction Models? *Economic Review* 84 (1999), 36–55. https://econpapers.repec.org/RePEc:fip:fedker:y:1999:i:qii:p:35-55:n:v.84no.2

[5] Milton Friedman. 1957. *A Theory of the Consumption Function* (1st ed.). Princeton University Press, Princeton, NJ.

[6] Gabriel Garber, Atif Mian, Jacopo Ponticelli, and Amir Sufi. 2018. *Household Debt and Recession in Brazil.* Working Paper 25170. National Bureau of Economic Research. https://doi.org/10.3386/w25170

[7] Yun Kim. 2016. Macroeconomic effects of household debt: An empirical analysis. *Review of Keynesian Economics* 4 (04 2016), 127–150. https://doi.org/10.4337/roke.2016.02.01

[8] Wenli Li. 2013. The economics of student loan borrowing and repayment. *Business Review* Q3 (2013), 1–10. https://ideas.repec.org/a/fip/fedpbr/y2013iq3p1-10.html

[9] Marco Jacopo Lombardi, Madhusudan Mohanty, and Ilhyock Shim. 2017. *The real effects of household debt in the short and long run.* BIS Working Papers 607. Bank for International Settlements. https://ideas.repec.org/p/bis/biswps/607.html

[10] Atif Mian and Amir Sufi. 2011. House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis. *American Economic Review* 101, 5 (August 2011), 2132–56. https://doi.org/10.1257/aer.101.5.2132

[11] Atif Mian and Amir Sufi. 2018. Finance and Business Cycles: The Credit-Driven Household Demand Channel. *Journal of Economic Perspectives* 32, 3 (August 2018), 31–58. https://doi.org/10.1257/jep.32.3.31

[12] Atif R Mian and Amir Sufi. 2010. *Household Leverage and the Recession of 2007 to 2009.* Working Paper 15896. National Bureau of Economic Research. https://doi.org/10.3386/w15896

[13] Atif R Mian, Amir Sufi, and Emil Verner. 2015. *Household Debt and Business Cycles Worldwide.* Working Paper 21581. National Bureau of Economic Research. https://doi.org/10.3386/w21581

[14] Rickard Nyman and Paul Ormerod. 2018. Understanding the Great Recession Using Machine Learning Algorithms. arXiv:econ.GN/2001.02115

[15] Anna Zabai. 2017. Household debt: recent developments and challenges. *BIS Quarterly Review* (December 2017). https://ideas.repec.org/a/bis/bisqtr/1712f.html

A APPENDIX

Table 1: Data Sets

Num	Description	Source
1	1 Month Treasury	https://fred.stlouisfed.org/series/GS1M
2	3 Month Treasury	https://fred.stlouisfed.org/series/GS3M
3	1 Year Treasury	https://fred.stlouisfed.org/series/GS10
4	10 Year Treasury	https://fred.stlouisfed.org/series/GS10
5	30 Year Treasury	https://fred.stlouisfed.org/series/GS30
6	10 Year Real Rates	https://fred.stlouisfed.org/series/FII10
7	Automobile Loans	https://fred.stlouisfed.org/series/CARACBW027SBOG
8	Auto Dealer Sales	https://fred.stlouisfed.org/series/MRTSSM4411USN
9	Consumer Price Index	https://fred.stlouisfed.org/series/CPIAUCSL
10	County Codes	https://data.bls.gov/cew/doc/titles/area/area$_{t}itles.htm$
11	Credit Card Rate	https://fred.stlouisfed.org/series/TERMCBCCINTNS
12	Employee Cost Index	https://data.bls.gov/cgi-bin/surveymost?bls
13	Fed Effective Funds Rate	https://fred.stlouisfed.org/series/DFF
14	German 10 Year Yield	https://fred.stlouisfed.org/series/IRLTLT01DEM156N
15	Household Debt	https://www.newyorkfed.org/medialibrary/media/research/ national_economy/householdcredit/pre2003_data.xlsx and https://www.newyorkfed.org/medialibrary/media/research/ national_economy/householdcredit/pre2003_data.xlsx
16	Household Debt to Income By County	https://www.federalreserve.gov/releases/z1/dataviz/household_debt/
17	Household Debt to Income By State	https://www.federalreserve.gov/releases/z1/dataviz/household_debt/
18	Household Debt to GDP	https://fred.stlouisfed.org/series/HDTGPDUSQ163N
19	Median Home Prices	https://fred.stlouisfed.org/series/MSPUS
20	Non Farm Employmment (NFE)	https://download.bls.gov/pub/time.series/ce/ ce.data.00a.TotalNonfarm.Employment
21	Produce Price Index	https://fred.stlouisfed.org/series/PPIACO
22	Rental Vacancy	https://fred.stlouisfed.org/series/RRVRUSQ156N
23	S&P/Case-Shiller Index	https://fred.stlouisfed.org/series/CSUSHPISA
24	State Codes	https://www.bls.gov/respondents/mwr/electronic-data-interchange/appendix-d-usps-state-abbreviations-and-fips-codes.htm
25	Student Loans	https://fred.stlouisfed.org/series/SLOAS
26	Total Employee Compensation	https://data.bls.gov/pdq/SurveyOutputServlet
27	Unemployment Rate	https://fred.stlouisfed.org/series/UNRATE/
28	Urban Consumer Rent	https://fred.stlouisfed.org/series/CUSR0000SAS2RS
29	U.S. GDP	https://fred.stlouisfed.org/series/GDP

Table 2: Correlation Report For All Input Variables

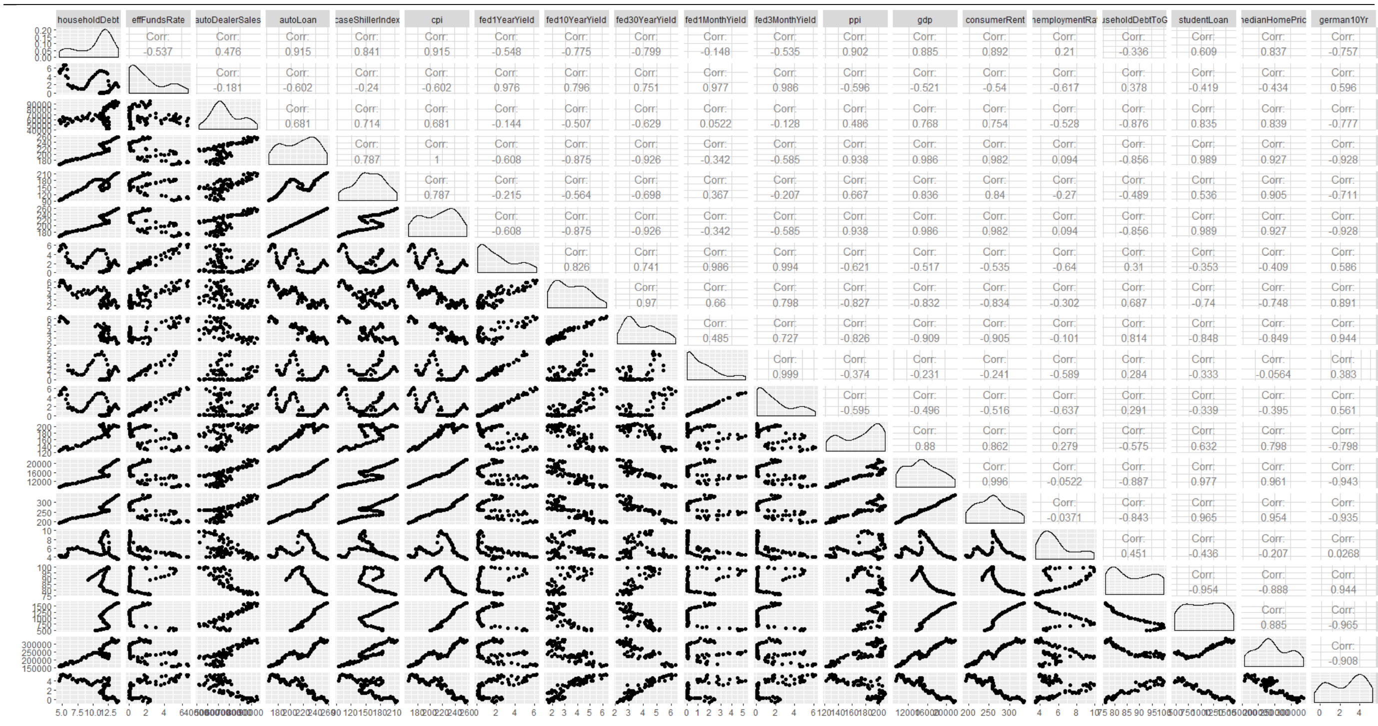

Table 3: Correlation Report For Final Input Variables

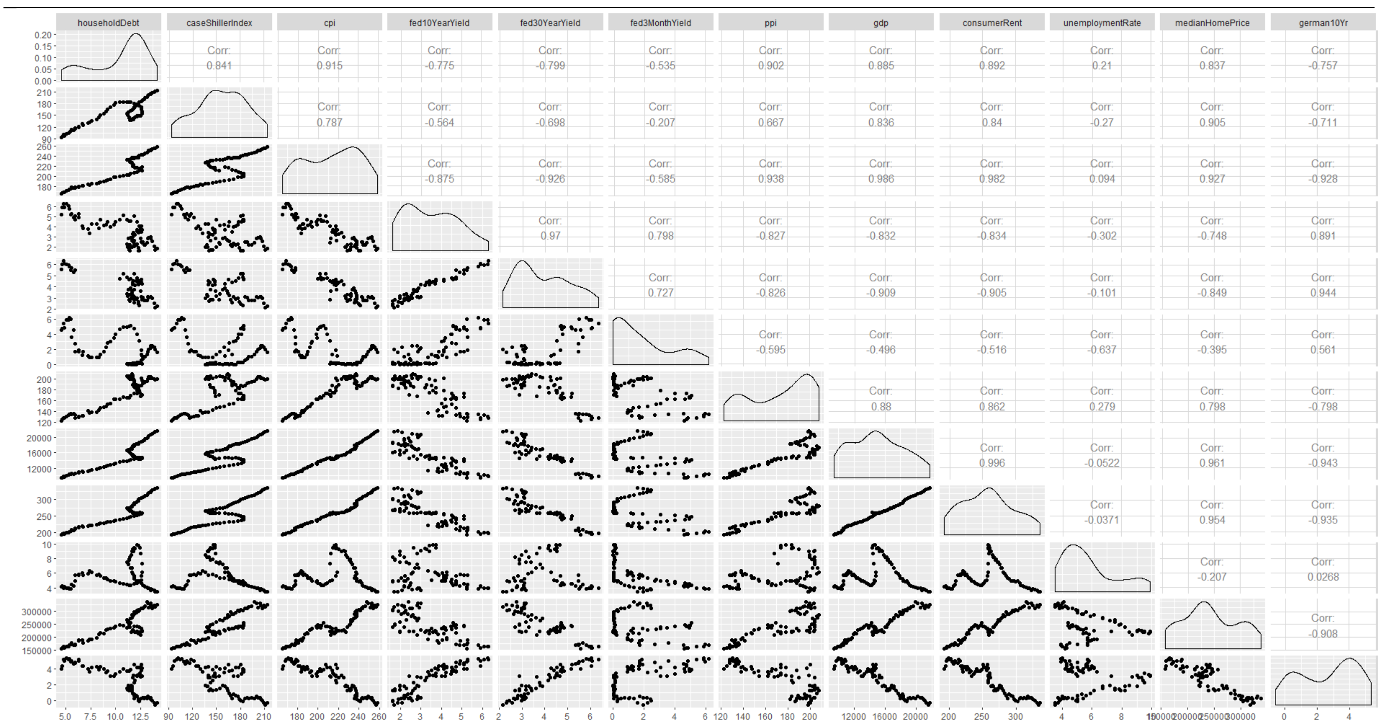

Table 4: Prediction vs Final Input Variables and Actual Household Debt

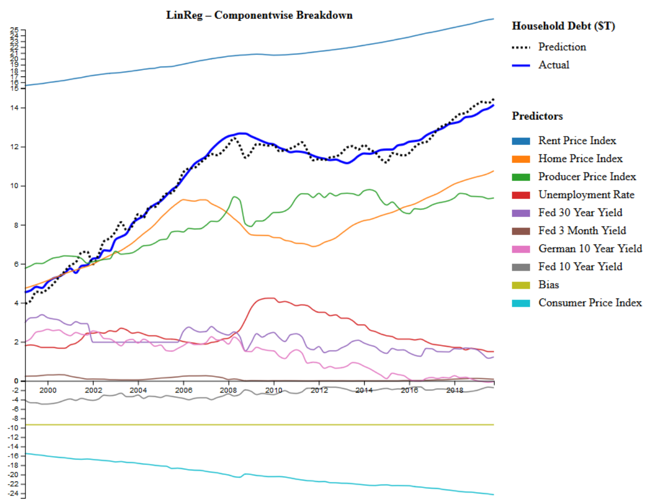

Table 5: Violin Plot of Debt On Quarterly Time Scale

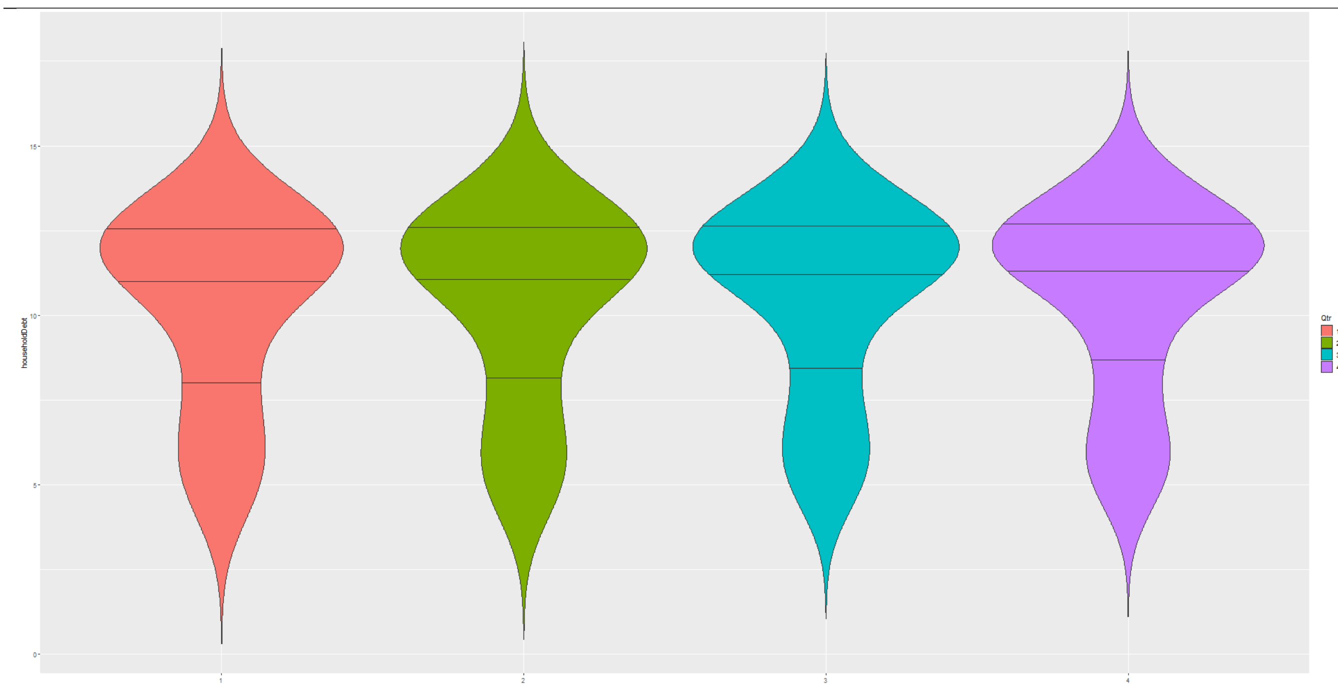

Table 6: Lasso Regression Plot of Mean Squared Error vs Num of Input Variables

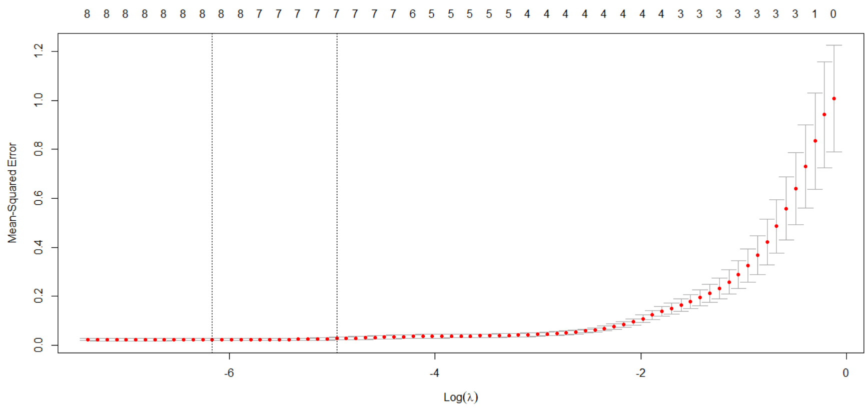

Table 7: Household Debt Auto Correlation

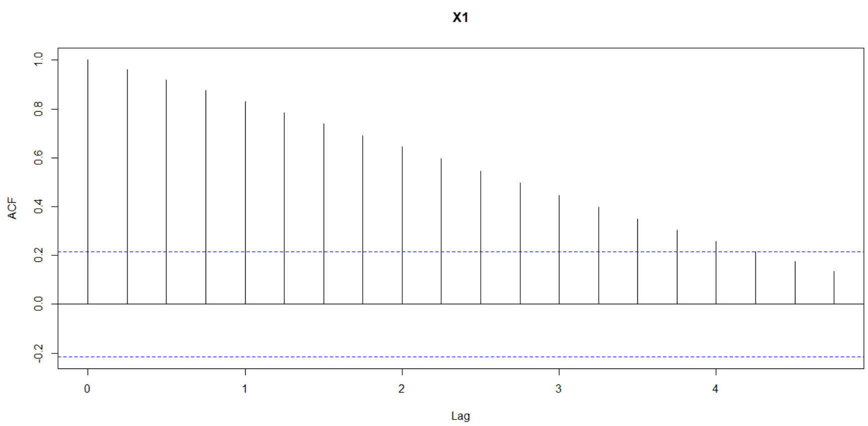

Table 8: Microsoft Azure Machine Learning Experiment

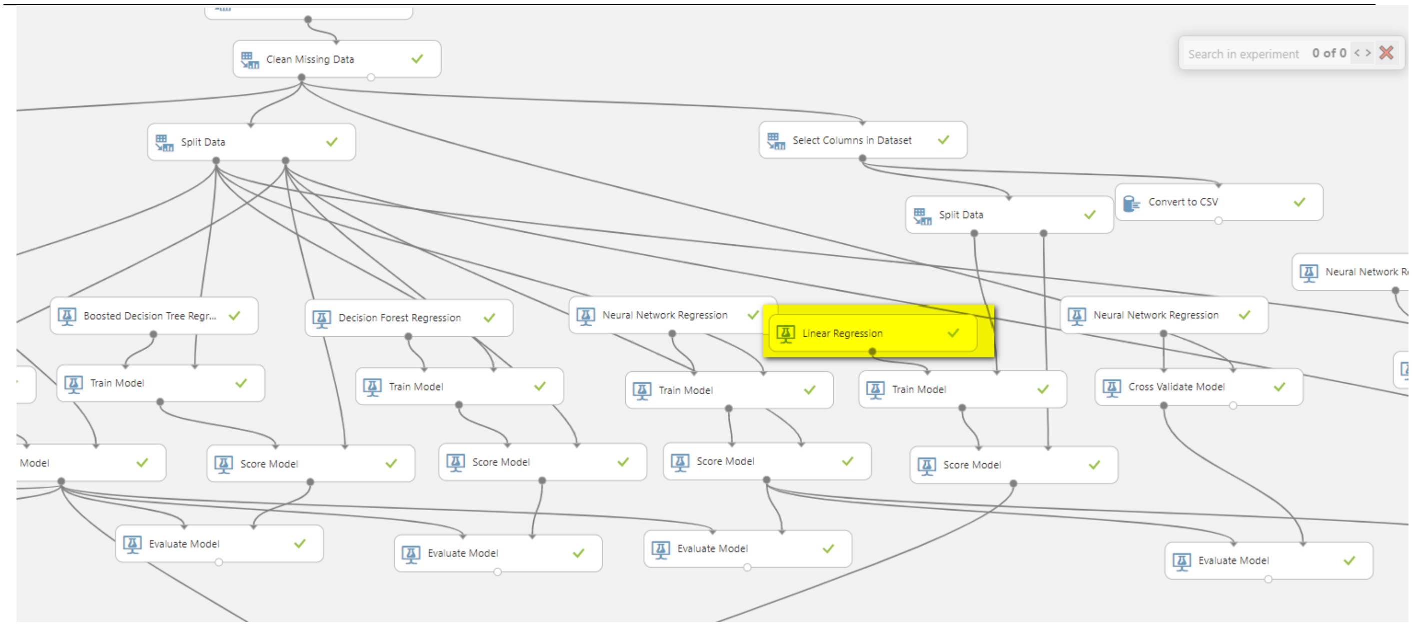

Table 9: Model Performance

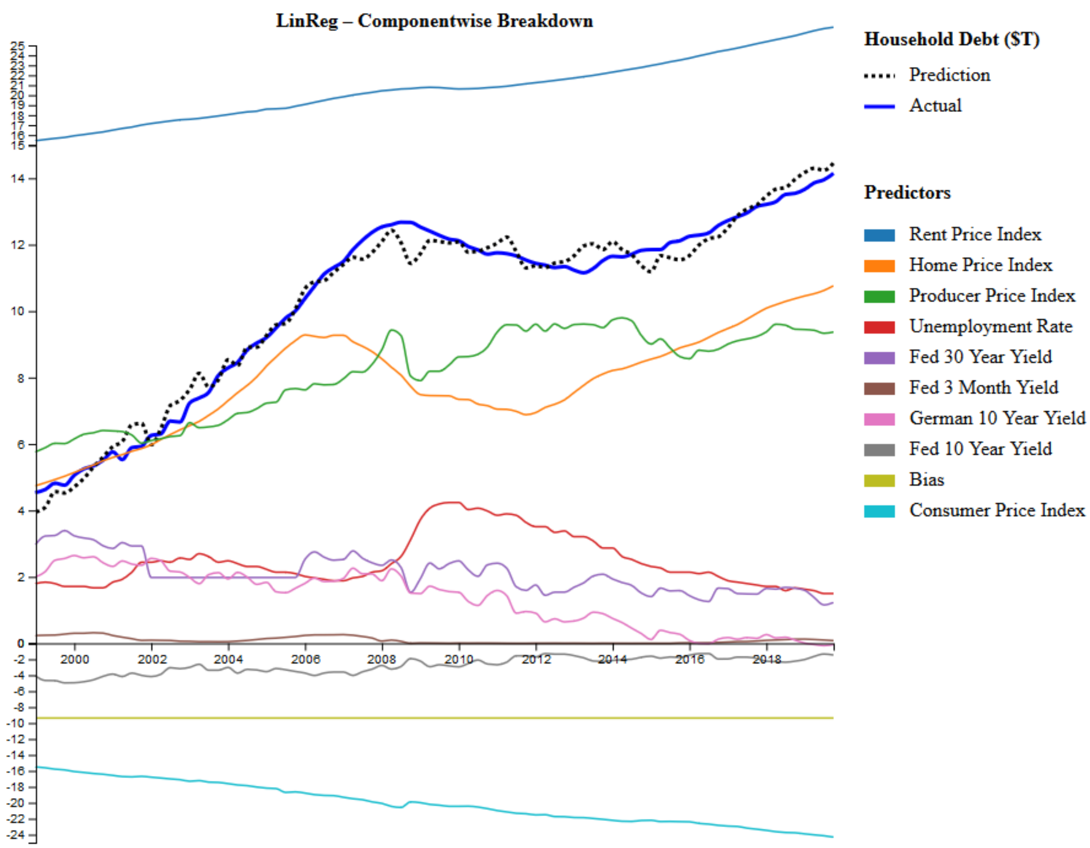

Table 10: Source Code

Source Folder	Description
root folder	The README.md file in this directory contains information about the project and team members. It also contains information about the project data sets
rcode	R source code used to generate plots to assist with variable selection. Please see the README.md file in this directory.
spark_scala_importer	Scala Apache Spark code to import data into SQLite, clean & transform data, and export to a CSV file used by Azure ML Studio. Please see the README.md file in this directory for setting up your IDE to compile the source code.
webapp	Angular and Bootstrap source code for web application. The Docker build files are also here. Please see the README.md file in this directory.