ASSIGNMENT 1

Group-2

ECON F342- APPLIED ECONOMETRICS

ANALYZING FACTORS CONTRIBUTING TO HOUSEHOLD INDEBTEDNESS AND FINANCIAL VULNERABILITY

Under Supervision of
Dr. Rishi Kumar Tiwari

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
HYDERABAD CAMPUS

GROUP MEMBERS

Serial No.	Name	ID Number
1	Abhinav Srivastava **(Leader)**	2021B3AA1608H
2	Anushka Singh	2021B3AA1484H
3	Anusha Makharia	2021B3A72681H
4	Ayushi Sharma	2021B3A71637H
5	Kinjal Vardia	2021B3AA2442H
6	Prakshaal Vora	2021B3A72967H
7	Rachit Vaghani	2021B3AA1457H
8	Shravani Daud	2021B3A72442H
9	Shubham Birla	2021B3A72965H
10	Suryashashank V.G.	2021B3AA0866H

ACKNOWLEDGEMENT

We express our sincere gratitude to Dr. Rishi Kumar for his unwavering support throughout the completion of our assignment titled "Analyzing Factors Contributing to Household Indebtedness and Financial Vulnerability." Dr. Rishi Kumar's guidance and availability were integral to the successful completion of our project. This assignment has proven intellectually stimulating, offering fresh perspectives and fostering a deeper comprehension of the intricate dynamics surrounding economic stability and individual financial well-being. Moreover, this undertaking has afforded us invaluable insights into the dichotomy between economic stability and vulnerability.

We sincerely appreciate the opportunity to learn under Dr. Rishi Kumar's mentorship. His profound guidance has played a pivotal role in shaping our insights and aiding us in navigating the complexities inherent in analyzing factors contributing to household indebtedness and financial vulnerability. We wish to emphasize the originality of our work and extend our heartfelt thanks to Dr. Rishi Kumar for providing a discerning and enriching learning experience.

ABSTRACT

This assignment examines the complex variables that affect household debt and financial vulnerability in a particular economic environment. Based mostly on large-scale data from credible organizations like the NSSO, our analysis uses regression diagnostics in the R programming language and the Ordinary Least Squares (OLS) approach.

Our main objective is to examine the effects of many factors, such as income, assets, education, caste , gender , state etc on the incidence of household debt. Understanding that there could be bias from missing data, we carefully added more variables to improve our model and reduce the overestimation of coefficients. Through several adjustments, we aim to develop a robust linear relationship between the independent and dependent variables, household indebtedness.

TABLE OF CONTENTS

1. Introduction

The study of household indebtedness and financial vulnerability has gained prominence, particularly in emerging economies like India. Understanding the complex dynamics of household finances is crucial for shaping overall economic stability and growth, emphasizing the need to investigate the factors influencing these aspects. This research aims to analyze household indebtedness and financial vulnerability in India thoroughly, employing econometric techniques on data from the All India Debt and Investment Survey (AIDIS).

Providing a concise background of the study and articulating the research questions is essential. This research seeks to identify the factors driving household indebtedness and financial vulnerability, exploring socio-economic and macroeconomic influences in the Indian context.

Several central inquiries guide this research:

- What factors contribute to household indebtedness in India?
- How do socio-economic factors such as income, employment, and education impact household financial vulnerability?
- To what degree do macroeconomic factors such as inflation, interest rates, and economic growth influence household indebtedness and financial vulnerability?

Addressing these questions is crucial for informing policymakers, researchers, and practitioners, as it can provide valuable insights into the underlying drivers of household financial stress and inform the development of targeted interventions. By understanding the factors shaping household indebtedness and vulnerability, stakeholders can devise more effective strategies to mitigate risks, enhance financial literacy, and strengthen economic resilience among Indian households.

2. Literature Review

Paper 1: An Empirical Analysis of Household Financial Vulnerability in India: Exploring the Role of Financial Knowledge, impulsivity, and Money Management Skills

Author: Kamakhya Nr Singh and Shruti Malik.

About the paper: The study examines the impact of household financial vulnerability in India, on low-income groups' financial stability during the COVID-19 epidemic. The study examines the relationship between financial vulnerability and impulsivity, money management skills, and financial literacy. The study aims to provide information on the factors leading to financial vulnerability in India through a nationwide household survey.

Methodology: Based on self-reported financial difficulties, perceptions of income shocks, and perceptions of expenditure shocks. The study proposes a Financial Vulnerability Index (FVI). Fractional regression examines how different behavioral and socioeconomic characteristics affect the financial vulnerability index.

Key Findings: Some factors can reduce financial vulnerability, including better money management skills. , less impulsive financial behavior, and increased financial education.
According to the study, vulnerable social groups should have special financial literacy programs that help them become better money managers, increase financial literacy, and deal with impulsivity. This helps them make more informed financial decisions and improve their financial well-being.

Conclusion: The study adds to our understanding of household financial vulnerability in India and provides recommendations for government agencies and financial regulators. The study emphasizes the importance of behavioral factors and financial education in reducing financial vulnerability and thus emphasizes the need for targeted interventions to improve the financial resilience of India's diverse population.

Paper 2: Class and Vulnerability to Debt in Rural India: A Statistical Overview

Author: Sandeep Kandikuppa

About the paper: In India, rural debt is a significant barrier to growth that affects different rural household classes differently. This research employs a class-based methodology to examine the subtleties of rural indebtedness using data from the All-India Debt and Investment Survey. The study demonstrates how socioeconomic and contextual vulnerabilities affect rural households of all socioeconomic groups' borrowing habits and debt loads.

Methodology: Focusing on 188,648 rural families in India, the study uses data from the All-India Debt and Investment Survey for the years 1992, 2003, and 2013. Motivated by earlier studies in the field, a four-class framework with roots in Marxist tradition is used to study rural indebtedness. Statistical techniques such as logistic regression are employed to investigate various outcomes connected to debt and indebtedness.

Key Findings:

- Wealthy rural households are more likely to take out larger loans, while asset-poor households are more likely to turn to unofficial lending sources.
- Compared to other groups, the wage worker class households are likelier to be too indebted.
- Compared to other classes, petty commodity producers have less debt and are less likely to be over-indebted.

Conclusion: The research highlights the intricacy of rural debt in India and the necessity for scholars, media, and policymakers to go beyond oversimplified accounts. This research highlights the need for a more nuanced understanding of rural indebtedness and the variables influencing it by considering how rural families experience debt depending on their socioeconomic level and contextual vulnerabilities.

<u>Paper 3</u>: The determinants of household indebtedness: household level evidence from Thailand

Author: P. Chotewattanakul, K. Sharpe, and S. Chand

About the paper: This study examines Thai household debt determinants using official surveys' of household data. Analyzing variables such as income stability, social class, access to formal credit sources, and the effect of different types of credit on debt development, the study aims to provide an overview of the factors influencing the debt of Thai households.

Methodology. : The study uses quantitative methods, including household ordinary least squares (OLS) estimation, to analyze the debt ratio and a two-part analysis of household over-indebtedness. Data from Thailand's Official Socio-Economic Household Survey, supplemented by the Bank of Thailand survey, are analyzed.

Key findings:

- Income stability, financial literacy, and access to formal credit sources are important factors affecting household debt in Thailand.
- Different loan types, such as consumer and agricultural loans, affect the probability of a household's objective over-indebtedness.
- Working household heads are better off in debt, but the share of over-indebtedness is higher, emphasizing age-related factors in household debt.

Conclusion: This study adds to understanding the determinants of household debt in Thailand and emphasizes the importance of income stability, financial literacy, and access to formal credit in developing household debt. The study distinguishes between subjective and objective over-indebtedness and provides valuable information for policymakers and financial institutions to address the problems of Thai household debt.

<u>Paper 4</u>: Household debt burden and financial vulnerability in Luxembourg

Author: Gaston Giordana and Michael Ziegelmeyer

About the paper: In analyzing family debt and financial vulnerability in Luxembourg, this research takes note of the central bank's worries over short mortgage rate fixing periods and rising debt-to-asset ratios. The study examines data from the Luxembourg Household Finance and Consumption Survey and finds mixed results, including a rise in median debt but a decrease in the percentage of families in debt. Identifying financially susceptible households highlights the significance of socio-economic elements in understanding household financial resilience. It also reveals relationships with age, income, and wealth.

Methodology: Using econometric methodologies, this paper uses data from the Luxembourg Household Finance and Consumption Survey (LU-HFCS) waves 2010 and 2014 to examine household debt burden and financial vulnerability in Luxembourg. It computes several metrics for indebted families, such as debt-to-income and debt-to-asset ratios. Financially vulnerable households are identified using predetermined thresholds for these indicators.

Key Findings: Younger, more populous, with dependent children, higher levels of education, and a greater chance of work are characteristics of indebted households. In 2014, the aggregate debt of families in debt rose by 27%, primarily as a result of mortgage debt. Debt burden measures are heavily influenced by age, net worth, and gross income, with higher net wealth correlated negatively with most debt burden measures.

Conclusion: 2014 saw a rise in the burden of debt for households, primarily due to mortgage loans. The ratios of outstanding loans-to-value and debt-to-income showed statistically significant rises. Lower costs associated with non-mortgage debt were the reason for the drop in the median debt service-to-income ratio. Based on predefined parameters, households that were financially insecure were selected. Future research on data-driven thresholds and including other macro situations in stress tests are also advised by the study.

<u>Paper 5</u>: Household indebtedness and financial fragility

Author: Tullio Jappelli, Marco Pagano, Marco Di Maggio

About the paper: The paper examines whether worldwide differences in household indebtedness lead to high debt levels and increased financial vulnerability. This is measured by households' susceptibility to insolvencies in response to lending and economic shocks. It also explores how institutional factors like information sharing, judicial efficiency, and bankruptcy regulations affect financial fragility.

Methodology: The paper employs a comprehensive methodology involving empirical analysis, regression models, and cross-country data. It compares the data across countries, and utilizes vector autoregressions to examine dynamic interactions. The sample is analyzed statistically, and regression modeling is done to explore the determinants of household indebtedness and financial vulnerability..

Key Findings: The paper shows that households are more likely to go bankrupt when they have more debt. Different countries have different levels of debt, and there is also a variation in the financial stability of households, but things are becoming more similar over time. When unexpected things happen, like moderate crises, many households might be unable to pay their debts, showing how important it is to have prominent rules in place. Lastly, stronger rules for creditors and better information sharing can reduce the chances of people defaulting on their debts.

Conclusion: The paper emphasizes the relationship between household indebtedness and financial vulnerability, supported by empirical evidence across countries. Institutional factors, such as creditor rights and information-sharing arrangements, significantly influence household debt levels and default rates. The study focuses on the importance of understanding the drivers of household defaults and the role of institutions in shaping the stability of household credit markets.

3. Data & Methodology

I. The dataset we use is the All India Debt and Instrument Survey (AIDIS), 2019, conducted by the National Statistical Office (NSO). It is the primary instrument source of data on various indicators of stock of assets, incidence of indebtedness, capital formation, and other indicators of rural/urban economy. These are used for planning policy formation and are input for further analytical studies by various Government organizations, academicians, researchers, and scholars.

II. Our model has **log(total_debt)** as the dependent variable with the independent variables log_monthly_exp, new_log_total_assets , education_factor , log(caste_factor) , gender_factor , State_factor , whether_holding_card , Household_size , and education dummy variables **(δ11-δ1_10)** , caste dummy variables **(δ21-δ25)** , state dummy variables **(δ31-δ65)**

$$
\begin{aligned}
\boldsymbol{log}(\boldsymbol{total\_debt} + \mathbf{1}) &= \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(log\ \_monthly\_exp) + \boldsymbol{\beta_2}(new\_log\ \_total\_assets) \\
&+ \boldsymbol{\beta_3}(Education\_factor\ ) + \boldsymbol{\beta_4} log\ (Caste\_factor\ ) + \boldsymbol{\beta_5}(Gender\_factor) \\
&+ \boldsymbol{\beta_6}(State\_factor) + \boldsymbol{\beta_7}(whether\_holding\_card) + \boldsymbol{\beta_8}(household\_size) \\
&+ (\delta_{11} + \delta_{12} + \delta_{13} + \delta_{14} + \delta_{15} + \delta_{16} + \delta_{17} + \delta_{18} + \delta_{19} + \delta_{1\_10}) + (\delta_{21} + \delta_{22} \\
&+ \delta_{23} + \delta_{24} + \delta_{25}) + (\delta_{31} + \delta_{32} \ldots + \delta_{65})
\end{aligned}
$$

iii. Justification of the variables chosen:

Variables	**Description**	**Justification**
log(total_debt+1)	The sum of all the cash loans, excluding balances due on overdraft accounts or cash credit limits sanctioned by a bank and Small loans taken for a short period without any security.(of the total household)	We'll look at the natural log of the dependent variable and see how it changes based on the coefficient of independent variables. This type of analysis is used when the relationship between variables is not linear, and taking the log helps to transform the data

		to fit a linear model better.
Caste_factor	Whether or not the household belongs to Scheduled Tribes, Scheduled Castes or Other Backward Classes will be indicated against this item in terms of the specified codes, which are: Scheduled Tribes - 1, Scheduled Castes - 2, Other Backward Classes - 3, other - 9	Some social groups face higher levels of indebtedness due to unequal access to resources, while others might have better financial stability. Researchers find a moderate negative correlation between household indebtedness and social status.
Gender_factor	The gender of each member will be recorded in this column in the code gender (male -1, female -2, transgender -3)	Choosing gender as an independent variable in the model allows us to investigate potential gender-based disparities in access to financial resources & income levels within households. Male-headed households generally have less debt accumulation due to societal privileges.
Education_factor	Highest level of education: Information regarding the level of general/technical/vocational educational level attained by the members of the household listed.	Including the highest education level can show whether having more education is linked to lower or higher debt, helping us see if education impacts how people manage their finances and borrow money. We expect this to have a negative correlation.
log_monthly _expenditure		We generally expect monthly expenditure to have a positive correlation with total household debt. This means that as monthly expenditure increases, we anticipate that total

		household debt will also increase. This helps us with debt management.
household_size	The total number of household members will be recorded against this item.	This variable facilitates exploring the impact of household size on debt levels, allowing for an analysis of how the number of individuals in a household influences borrowing behavior and providing insights into the dynamics of household finances and expenditure patterns.
whether_holding_card	whether holding a credit/debit card (yes-1, no-2)	This variable allows for investigating the relationship between household debt and the ownership of credit or debit cards, providing insights into how access to these financial instruments influences borrowing behavior.
State_factor	Which state does the household belong to?	Including the "states" variable as an independent variable allows for examining how variations in geographic location impact household debt, providing insights into regional disparities and the influence of local economic, policy, and demographic factors on borrowing behavior.
new_log_total_assets	Log of total assets combined, including cash and non-cash assets like shares and similar instruments.	Understanding how this relates to "total debt" helps us see if wealth affects debt. If assets go up, debt might go down, showing better

		financial health. It's like seeing how much money you have versus how much you owe.
δ11-δ1_10	Education dummy variables	Different levels of education within households are important because education influences financial habits. We expect higher education levels to correlate with lower debt due to better financial skills and higher incomes.
δ21-δ25	Caste dummy variables	These variables helps us understand if certain caste groups tend to have more or less debt than others. Knowing this can guide efforts to reduce financial inequalities and promote fairness.
δ31-δ65	States dummy variables	These dummy variables allow us to account for differences between states, giving insights into regional disparities & informing policymaking. It also helps to consider factors unique to each state that might affect household debt.

The checks we do to ensure that the data satisfies all the assumptions:

1. **<u>Omitted Variable Test</u>**

   Command: resettest (formula, power = 2:3, type = “regressor”, data = name_dataset, ...)

   Theory: The Ramsey Regression Equation Specification Error Test, or RESET, determines whether any variables in the regression model have been left out. It investigates if non-linear combinations of the explanatory variables—their powers—help explain the dependent variable. Since this method is a "lmtest" library component, the test must first be installed and loaded.

   Hypothesis Set: **Null Hypothesis**: The model has no omitted variables.
   Alternate Hypothesis: The model has omitted variables.

2. **<u>Shapiro-Wilk Test for Normality</u>**

   Command: shapiro.test (name_variable)

   Theory: The Shapiro-Wilk test is essentially a goodness-of-fit test. Otherwise, it evaluates how well the sample data conforms to a normal distribution.

   Hypothesis Set: **Null Hypothesis**: The population is normally distributed.
   Alternate Hypothesis: The population is not normally distributed.

3. **<u>Breusch-Pagan Test for Heteroscedasticity</u>**

   Command: ols_test_breusch_pagan (formula, data = name_dataset)

   Theory: This assesses if the model exhibits heteroskedasticity or whether the values of the independent variables have any bearing on the variance of the errors arising from a regression. Since this function requires the "olsrr" library to be installed and loaded, it must be done before the test can be conducted.

Hypothesis Set: **<u>Null Hypothesis</u>**: The model is homoscedastic.

<u>Alternate Hypothesis:</u> The model is heteroscedastic.

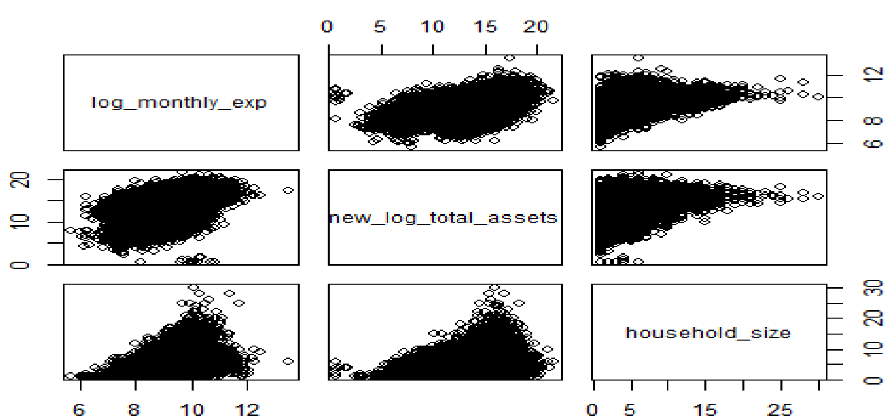
4. <u>Test for Multicollinearity</u>

Command: ols_vif_tol (formula, data = name_dataset)

Theory: The tolerance measurement determines the percentage of the variance in the independent variable that cannot be accounted for by the other independent variables. Therefore, the other independent variables can sufficiently explain an increase or fall in the value of the relevant independent variable if the tolerance is lower. Because of this, multicollinearity might exist, given the correlation between the variables.

The inflation included in the independent variable's coefficient is measured by the Variance Inflation Factor (VIF), which considers collinearities between the various independent variables. Multicollinearity does not exist if the VIF is 1, as the presence of the other factors does not cause the regression coefficient to increase. To put it simply, VIF is the antithesis of tolerance. Therefore, a lower VIF value suggests the absence of multicollinearity. A VIF close to 1 indicates little correlation among the predictors, suggesting that multicollinearity is not a significant issue.

<u>Graph Matrix</u>



In the graph matrix, we plot every quantitative variable against each other to get the final plot. We interpret this graph with the rows showing the relationship between the variables. We see an exponential decaying positive relation between log_monthly_exp and the household_size. This shows that after a point, even after the household size increases, the monthly expense plateaus. Additionally, we also see that there is a similar positive relationship between the log_total_assets and household_size

4. Results and Discussion

Regression diagnostics:

- **Omitted Variable test**

```
> reset_test = resettest(model,power=2)
> print(reset_test)

        RESET test

data:  model
RESET = 46.184, df1 = 1, df2 = 113348, p-value = 1.082e-11
```

We reject the null hypothesis at 5% level of significance as the p-value is less than 0.05, and conclude that the model contains omitted variables.

- **Heteroskedasticity**

```
> bp_test <- bptest(model)
> print(bp_test)

        studentized Breusch-Pagan test

data:  model
BP = 5967, df = 54, p-value < 2.2e-16
```

We reject the null hypothesis at a 5% significance level (p-value > 0.05) and conclude that the model is heteroskedastic.

- **Normality Test**

```
> sample_data <- HouseholdMergedFinal$new_log_total_debt[sample(seq_along(HouseholdMergedFinal$new_log_total_debt), 5000)]
>
> # Perform Shapiro-Wilk test on the sampled data
> shapiro.test(sample_data)

        Shapiro-Wilk normality test

data:  sample_data
W = 0.77592, p-value < 2.2e-16
```

We reject the null hypothesis at a 5% significance level (p-value < 0.05) and conclude that the population is not normally distributed.

- **Multicollinearity**

```
> vif_values <- car::vif(model)
> print(vif_values)
                         GVIF Df GVIF^(1/(2*Df))
log_monthly_exp      2.213973  1        1.487943
new_log_total_assets 1.243063  1        1.114928
Education_factor     1.699010 10        1.026857
Caste_factor         1.860177  3        1.108985
Gender_factor        1.115215  2        1.027637
State_factor         2.653456 35        1.014039
whether_holding_card 1.388141  1        1.178194
household_size       1.740017  1        1.319097
```

VIF (Variance Inflation Factor) is an index that identifies the correlation between independent variables. If VIF>10, we infer multicollinearity exists in the model. The mean VIF of the model comes out at less than 10. This shows there is no multicollinearity in the model.

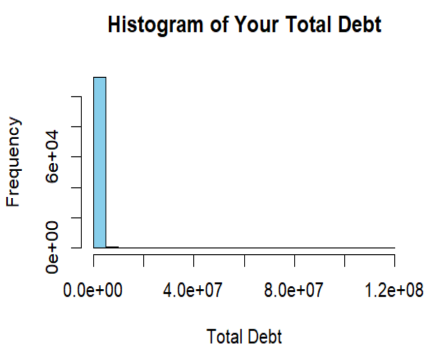
Summary Statistics of the Variables Use

1. Total Debt

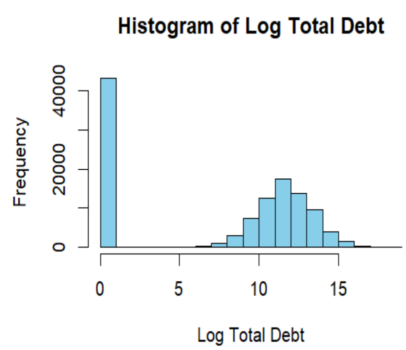
```
> describe(HouseholdMergedFinal$total_debt)
   vars      n     mean      sd median  trimmed      mad min       max     range  skew kurtosis      se
X1    1 113750 295651.5 1311278  28250 90763.07 41883.45   0 115200000 115200000 28.06  1620.21 3887.94
```

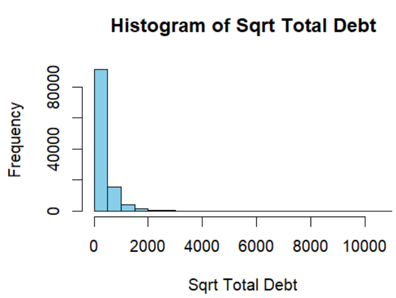
Total debt is skewed to the right, so we take a log transformation, making it log(total_debt + 1) If the data is skewed or not normally distributed, taking the values' logarithm can help make the distribution more symmetric. This is particularly useful for variables with a wide range of values.

Total Debt Graph



Log Total Debt Graph





Taking Sqrt debt significantly improves the model's **R^2** instead of taking log as sqrt compresses the distribution.

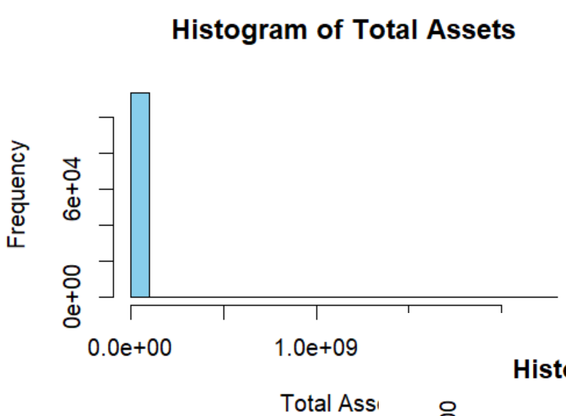
2. Total Assets

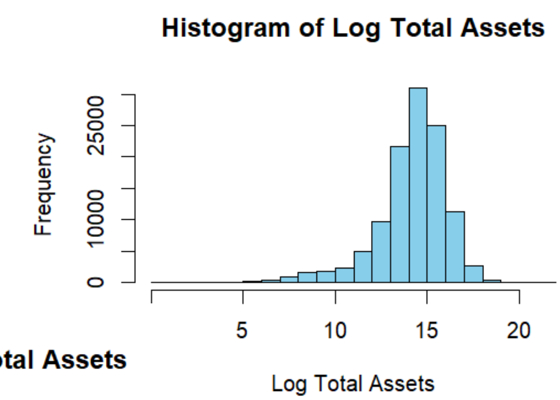
Total assets is skewed to the right, so we take a log transformation, making it log(total_assets+1). If the data is skewed or not normally distributed, taking the values' logarithm can help make the distribution more symmetric. This is particularly useful for variables with a wide range of values. We are also taking square root transformation, similar to the log transformation, which can stabilize variance and make the distribution more normal.

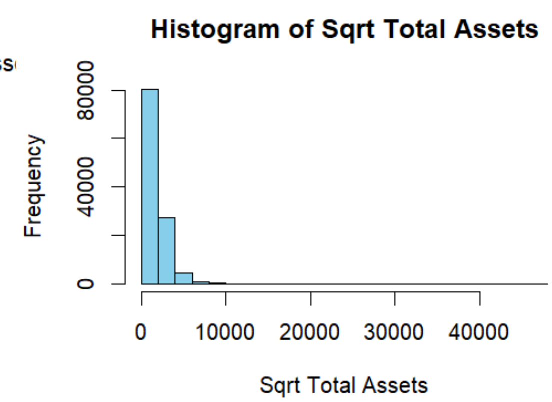
```
> describe(HouseholdMergedFinal$new_log_total_assets)
   vars      n  mean  sd median trimmed  mad  min   max range  skew kurtosis   se
X1    1 113750 14.17 1.9  14.44   14.36 1.47 0.69 21.55 20.86 -1.27     2.92 0.01
```

Kurtosis=2.92 suggests that the distribution is moderately peaked compared to normal distribution but not as extreme as it would be if the kurtosis were significantly higher.

Total Assets Graph **Log Assets Graph**

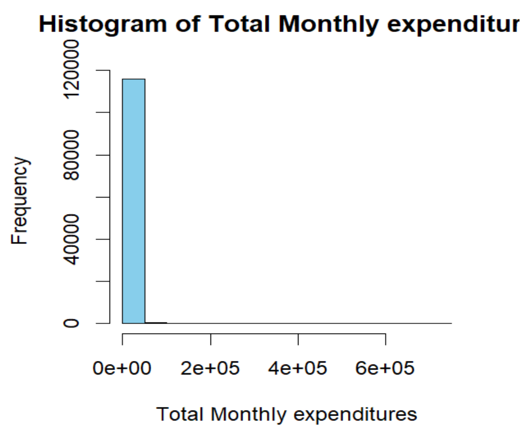




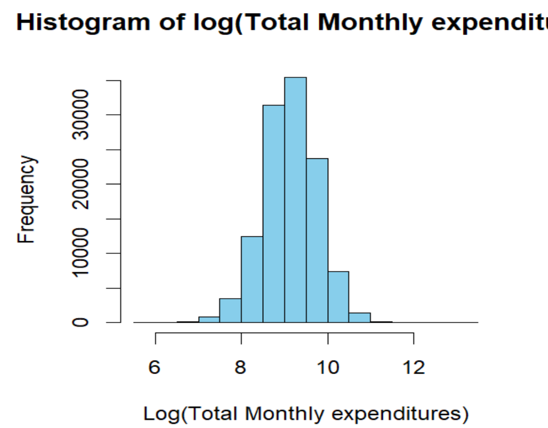


3. Total Monthly Expenditure

Monthly Total Expenditure



Log(Monthly Total Expenditure)



The histogram plot indicates that the distribution of monthly expenditures is skewed to the right. To address this skewness and make the distribution more symmetric, a log transformation is applied to the monthly expenditure values, making it log(monthly expenditure+1). The right-skewed distribution of monthly expenditures suggests relatively more households with lower expenditure values than higher ones. This transformation is particularly beneficial when dealing with variables that exhibit a wide range of values and skewed distributions.

4. Highest Level of Education

```
> summary(HouseholdMergedFinal$Education_factor)
                                 Illiterate                               Below Primary
                                      29864                                        9098
Diploma/Certificate (Graduation and above)     Diploma/Certificate (Higher Secondary)
                                        732                                         824
               Diploma/Certificate (Secondary)                                  Graduate
                                        543                                       10486
                           Higher Secondary                     Postgraduate and above
                                       9954                                        3342
                                    Primary                                   Secondary
                                      15321                                       15674
                       Upper Primary/Middle                                        NA's
                                      17910                                        2713
```

We see that a maximum of STs, SCs, and OBCs are illiterate, while only a tiny fraction of these communities have attained a diploma or certificate equivalent to secondary education. The prevalence of illiteracy among STs, SCs, and OBCs underscores the enduring educational obstacles encountered by marginalized communities. These challenges impede their access to high-quality education, perpetuates a cycle of limited opportunities and social inequity. Addressing this disparity demands focused efforts to enhance educational accessibility and quality within these communities, fostering inclusive development and empowering individuals to break free from the constraints of illiteracy.

Interpretation of Model 1

```
Call:
lm(formula = new_log_total_debt ~ log_monthly_exp + new_log_total_assets +
    Education_factor + Caste_factor + Gender_factor + State_factor +
    whether_holding_card + household_size, data = HouseholdMergedFinal)

Residuals:
    Min      1Q  Median      3Q     Max
-15.564  -5.942   2.534   4.449  11.974

Coefficients:
                                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                                        -8.845095   0.331833 -26.655  < 2e-16 ***
log_monthly_exp                                     0.809400   0.038138  21.223  < 2e-16 ***
new_log_total_assets                                0.639753   0.009765  65.514  < 2e-16 ***
Education_factorBelow Primary                       0.253019   0.066955   3.779 0.000158 ***
Education_factorDiploma/Certificate (Graduation and above) -1.004863   0.207645  -4.839 1.30e-06 ***
Education_factorDiploma/Certificate (Higher Secondary)    -0.437963   0.196575  -2.228 0.025884 *
Education_factorDiploma/Certificate (Secondary)    -0.532503   0.240310  -2.216 0.026701 *
Education_factorGraduate                           -0.842849   0.071297 -11.822  < 2e-16 ***
Education_factorHigher Secondary                   -0.386578   0.068865  -5.614 1.99e-08 ***
Education_factorPostgraduate and above             -0.862613   0.107664  -8.012 1.14e-15 ***
Education_factorPrimary                             0.270357   0.056385   4.795 1.63e-06 ***
Education_factorSecondary                          -0.188879   0.059062  -3.198 0.001384 **
Education_factorUpper Primary/Middle                0.271950   0.055257   4.922 8.60e-07 ***
Caste_factorOther                                   0.353553   0.062687   5.640 1.70e-08 ***
Caste_factorOther Backward Class (OBC)              0.669242   0.059015  11.340  < 2e-16 ***
Caste_factorScheduled Caste (SC)                    0.842255   0.066672  12.633  < 2e-16 ***
Gender_factorFemale                                -0.872070   0.051046 -17.084  < 2e-16 ***
Gender_factorTransgender                           -1.331964   1.169210  -1.139 0.254622
State_factorAndaman and Nicobar Islands            -2.953438   0.359945  -8.205 2.32e-16 ***
State_factorAndhra Pradesh                          0.314016   0.132865   2.363 0.018109 *
State_factorArunachal Pradesh                      -4.378075   0.194719 -22.484  < 2e-16 ***
State_factorAssam                                  -2.458711   0.139118 -17.674  < 2e-16 ***
State_factorBihar                                  -2.335434   0.121977 -19.147  < 2e-16 ***
State_factorChandigarh                             -2.612435   0.423157  -6.174 6.69e-10 ***
State_factorChhattisgarh                           -2.302405   0.156479 -14.714  < 2e-16 ***
State_factorDadra and Nagar Haveli                 -1.282779   0.415707  -3.086 0.002031 **
State_factorDaman and Diu                          -0.999011   0.413911  -2.414 0.015798 *
State_factorDelhi                                  -4.825361   0.173124 -27.872  < 2e-16 ***
State_factorGoa                                    -2.268659   0.377354  -6.012 1.84e-09 ***
State_factorGujarat                                -2.187756   0.129601 -16.881  < 2e-16 ***
State_factorHaryana                                -2.502865   0.157952 -15.846  < 2e-16 ***
State_factorHimachal Pradesh                       -2.069366   0.199878 -10.353  < 2e-16 ***
State_factorJammu and Kashmir                      -2.513690   0.173037 -14.527  < 2e-16 ***
State_factorJharkhand                              -2.123027   0.147331 -14.410  < 2e-16 ***
State_factorKarnataka                              -1.290511   0.126460 -10.205  < 2e-16 ***
State_factorKerala                                 -0.659509   0.139659  -4.722 2.33e-06 ***
State_factorLakshadweep                            -2.621451   0.472613  -5.547 2.92e-08 ***
State_factorMadhya Pradesh                         -1.428927   0.126009 -11.340  < 2e-16 ***
State_factorMaharashtra                            -1.682321   0.117322 -14.339  < 2e-16 ***
State_factorManipur                                -2.930816   0.155698 -18.824  < 2e-16 ***
State_factorMeghalaya                              -4.450919   0.187033 -23.798  < 2e-16 ***
State_factorMizoram                                -2.857161   0.196130 -14.568  < 2e-16 ***
State_factorNagaland                               -4.021815   0.211743 -18.994  < 2e-16 ***
State_factorOdisha                                 -0.721370   0.136336  -5.291 1.22e-07 ***
State_factorPuducherry                             -1.235715   0.313750  -3.939 8.20e-05 ***
State_factorPunjab                                 -1.803223   0.149923 -12.028  < 2e-16 ***
State_factorRajasthan                              -1.545033   0.125704 -12.291  < 2e-16 ***
State_factorSikkim                                 -2.711995   0.216644 -12.518  < 2e-16 ***
State_factorTamil Nadu                             -1.083571   0.122977  -8.811  < 2e-16 ***
State_factorTripura                                -1.630608   0.155592 -10.480  < 2e-16 ***
State_factorUttar Pradesh                          -2.022558   0.114987 -17.589  < 2e-16 ***
State_factorUttarakhand                            -2.472097   0.194474 -12.712  < 2e-16 ***
State_factorWest Bengal                            -1.491376   0.121349 -12.290  < 2e-16 ***
whether_holding_card                                0.308493   0.038351   8.044 8.77e-16 ***
household_size                                      0.216150   0.009954  21.715  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.481 on 113349 degrees of freedom
  (3057 observations deleted due to missingness)
Multiple R-squared:  0.1103,	Adjusted R-squared:  0.1099
F-statistic: 260.2 on 54 and 113349 DF,  p-value: < 2.2e-16
```

Variable	Estimate	Interpretation
log_monthly_exp(***)	0.809	The coefficient of 0.809 suggests that for every 1% increase in monthly expenses, total debt tends to increase by about 0.809%. This indicates a positive and proportional relationship between monthly expenses and total debt.
log_total_assets(***)	0.639	The coefficient of 0.639 suggests that for every 1% increase in total assets, total debt tends to increase by about 0.639%. This indicates a positive relationship between total assets and total debt.
(Education_Factor)i	-	Except Below Primary and Primary all the other estimates are negative indicating that higher level of education leads to lesser debt.
(State)i	-	Southern states are showing a trend to have a higher estimated value than northern states indicating that households in southern states have a higher total debt. Telangana has a +ve value.
Female	-0.87	Compared to the base variable(male), the probability of females taking loans is 87 percent less. After taking into consideration other variables included in the regression model, this coefficient indicates that, generally speaking, women are linked to lower loan amounts or a lower likelihood of taking out loans than men.
Whether Holding Credit/Debit	0.308	We see that households in

Card		which the head of the family has access to a credit or debit card/access to financial institutions take more loans
Household Size	0.216	The coefficient of 0.216 suggests that if the household size increases by 1 unit, the debt increases by 21.6%, ceteris paribus. This positive coefficient suggests a positive association between household size and total debt. In other words, larger households tend to have higher total debt levels, assuming all else remains constant.
(Caste)i	-	We can see that SC households tend to have the highest predicted values, followed by OBC households and then households belonging to other caste categories.

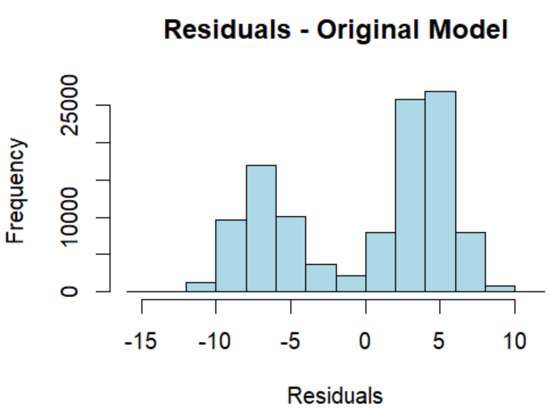
Remedial Measures Taken

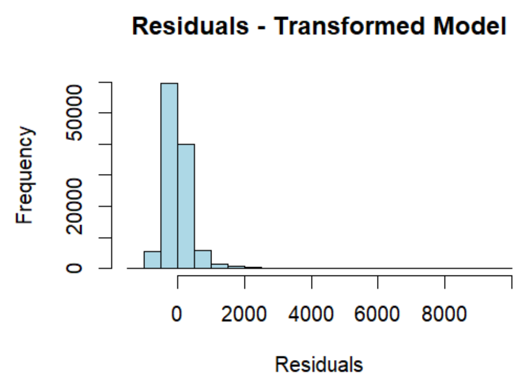
As the residuals were not normally distributed for our original model, our homoscedasticity assumption was violated, so we tried some transformations

i) log_total_debt was changed to sqrt_total_debt

ii) log_total_assets was changed to sqrt_total debt

This helped improve the distribution of residuals to become more like a normal distribution.





New Transformed Model Equation

model_sqrt<-lm(sqrt_total_debt_transformed~ log_monthly_exp + log_total_assets_transformed + Education_factor + Caste_factor + Gender_factor+ State_factor + whether_holding_card + household_size , data = HouseholdMergedFinal)

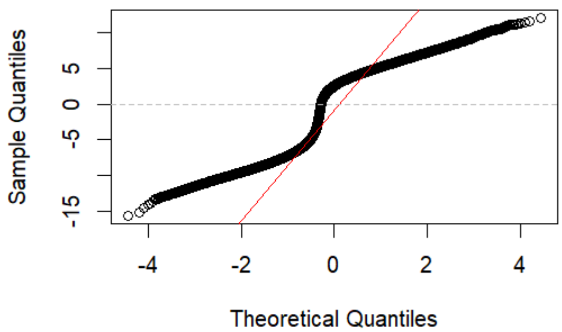
Original

```
Residual standard error: 408.4 on 113349 degrees of freedom
  (3057 observations deleted due to missingness)
Multiple R-squared:  0.1869,    Adjusted R-squared:  0.1865
F-statistic: 482.5 on 54 and 113349 DF,  p-value: < 2.2e-16
```

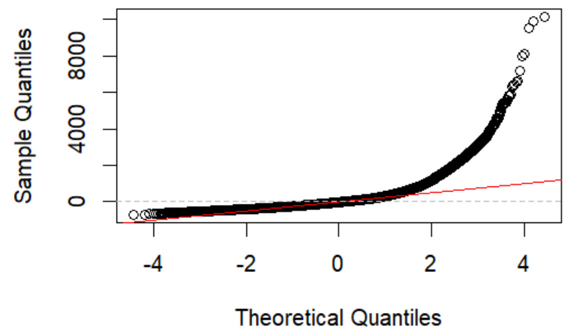
New Transformed Model

```
Residual standard error: 407.4 on 113349 degrees of freedom
  (3057 observations deleted due to missingness)
Multiple R-squared:  0.1911,	Adjusted R-squared:  0.1907
F-statistic: 495.8 on 54 and 113349 DF,  p-value: < 2.2e-16
```

Original Q-Q Plot



Transformed Q-Q Plot



- In the Q-Q plot analysis for the new model, we identified a pronounced S-shaped deviation in the higher quintiles of the residual distribution. This observation indicates a departure from the normal distribution assumption, particularly in the tails of the data.

5. Conclusion

In our study on household debt and financial fragility in India, we explored the complex interactions between several socio-economic determinants and their effects on households' financial well-being. Using a dataset from the National Statistical Office (NSO) All India Debt and Investment Survey (AIDIS) 2019, we applied robust regression analysis to identify the complex link between several factors and the prevalence of household debt.

Several conclusions emerged from our investigation. First, we discovered a positive and proportionate relationship between monthly spending and overall debt levels, suggesting that as monthly spending increases, so does a household's likelihood of accruing debt. Similarly, larger household debt levels were correlated with an increase in total assets, indicating the importance of wealth creation in indicating borrowing behavior.

Additionally, our study shed light on the relationship between household debt and sociodemographic characteristics. The findings indicate a positive correlation between household size and total debt, implying that bigger families often incur higher debt levels, maybe due to more financial commitments and spending demands. Another important factor surfaced was gender dynamics, with female-headed families showing lower debt levels than male-headed households.

Furthermore, our data showed that household debt levels varied by area, with some states showing greater debt levels than others. This emphasizes the importance of considering regional economic, social, and policy aspects to comprehend borrowing patterns and financial fragility.

To sum up, our research highlights the complex nature of family debt and advances our understanding of household finances in India. By clarifying the factors contributing to financial vulnerability and geographic differences, our research can help stakeholders and policymakers develop focused interventions to increase financial literacy, reduce risks, and strengthen economic resilience among Indian families.

6. References

1. K. N. Singh, Shruti Malik (2022). An empirical analysis on household financial vulnerability in India: exploring the role of financial knowledge, impulsivity, and money management skills. Link.
2. Sandeep Kandikuppa (2021).Class and Vulnerability to Debt in Rural India: A Statistical Overview. Link.
3. Chotewattanakul, Pasit (2019). The determinants of household indebtedness: household level evidence from Thailand. Link.
4. Gaston Giordana and Michael Ziegelmeyer (2017). Household debt burden and financial vulnerability in Luxembourg. Link.
5. Tullio Jappelli, Marco Pagano, Marco Di Maggio (2008). Households' Indebtedness and Financial Fragility. Link.