Understanding Household Debt Dynamics in Sweden: A Comprehensive Analysis

G.H. AMILA HETTIARACHCHI
STATISTICAL LEARNING – COURSE COORDINATOR: MOUDUD ALAM (SENIOR LECTURER STATISTICS)
APRIL 29, 2024

Abstract

In this study, the relationship between household debt and inflation in Sweden is investigated, with a focus on the impact of the Consumer Price Index (CPI). Historical data spanning from 1980 to 2009 is analyzed using rigorous statistical methods. The methodology involves examining various economic indicators and demographic variables to understand Sweden's economic landscape. The dataset is preprocessed to ensure suitability for analysis.

An inverse relationship between CPI and household debt is observed, although CPI's influence is not statistically significant. Correlation analysis highlights significant correlations between household debt and Real GDP per capita. Multicollinearity analysis identifies Real GDP per capita as a key predictor. By regenerating the model without highly correlated variables, improvements are achieved.

The linear regression model demonstrates strong predictive accuracy, explaining approximately 94.6% of household debt variability. Lasso regression emerges as the preferred model, balancing prediction accuracy and generalization ability effectively. For the year 2021, predictions of household debt using Lasso and Ridge models closely align with actual values.

Overall, this study provides valuable insights into household debt dynamics in Sweden, offering a framework for economists to understand and predict trends more effectively.

Table of Contents

Introduction:

Sweden witnessed a significant milestone in March 2021, as household debt soared to a record high, constituting 97.8% of the nation's Nominal GDP. This surge in debt prompts a critical inquiry into the potential impact of inflation, particularly as measured by the Consumer Price Index (CPI), on this phenomenon. Such understanding is pivotal for policymakers and economists alike, as it informs the formulation of effective strategies to uphold economic stability (CEIC Data, 2021). Thus, this paper aims to explore the nexus between household debt and inflation in Sweden, drawing insights from historical data and employing rigorous statistical analysis.

Methodology

Research Question

The research questions addressed in this study are as follows:

1. Does inflation, as measured by the Consumer Price Index (CPI), have a significant influence on the increase in household debt in Sweden?

2. How accurately can recent household debt in Sweden be predicted using historical data, considering the current level is approximately 87% of Nominal GDP?

Understanding the data

A variety of factors were examined in the dataset to gain insight into the economic landscape of Sweden between 1980 and 2009. Firstly, the "year" variable was utilized as a chronological reference point, delineating the temporal scope of the analysis. Real economic metrics, such as "rgd" (Real GDP per capita) and its growth rate ("growth"), were utilized to glean insights into the country's economic performance over time, sourced from the Penn World Tables 7.0.

Demographic variables, including "pop" (Population) and "school" (Average years of schooling), were employed to illuminate societal context, reflecting population trends and educational attainment levels derived from Barro & Lee (2000).

Furthermore, the "cpi" (Consumer Price Index) variable was employed to ascertain annual changes in consumer prices, aiding in the understanding of inflationary trends, sourced from the World Bank World Development Indicators. The "dep" (Total dependency ratio) variable was utilized to gain insights into demographic structure, measuring the proportion of dependent individuals relative to the working-age population, also sourced from the World Bank World Development Indicators. Additionally, the binary "crisis" variable was utilized to denote the occurrence of banking crises, assisting in the assessment of economic stability, sourced from Carmen M Reinhart's website.

Of relevance to the research inquiry was the "debthhld" variable, which indicated household and non-profit institutions serving households (NPISH) liabilities as a percentage of GDP, sourced from OECD and national sources. This variable served as a focal point for investigating the dynamics of household debt and its potential determinants within the Swedish economy. Collectively, these variables offered a comprehensive understanding of the economic, demographic, and financial landscape of Sweden during the specified period, laying the foundation for rigorous analysis and interpretation. The following Data dictionary is used throughout the research.

- debthhld: Household Debt
- cpi: Consumer Price Index (CPI)
- rgd: Real GDP per capita
- growth: Real GDP per capita growth
- pop: Population
- school: Average years of schooling
- dep: Dependency Ratio
- crisis: Banking Crisis

Data Preprocessing

The given dataset contains no missing values. Consequently, there is not much preprocessing required for the dataset. However, the data has been converted to numeric values, as necessitated for the linear regression analysis.

Method of analysis

The dataset was examined to explore the relationship between household debt and the Consumer Price Index (CPI). A scatter plot was generated to visually assess the potential linear connection between household debt and CPI. Here, CPI was treated as the independent variable, while household debt served as the dependent variable. After visualizing the data, linear regression analysis was applied to quantify the strength of the relationship and ascertain the slope.

Additionally, a correlation analysis was conducted to unveil any other factors influencing household debt beyond CPI. This analysis aimed to identify additional variables that could potentially explain variations in household debt.

Subsequently, the analysis expanded to forecast household debt for recent years. The goal was to develop and validate a predictive model capable of accurately estimating household debt. The accuracy of the model in predicting household debt was thoroughly evaluated to ensure its reliability for future predictions.

Data Analysis

During the data analysis phase, meaningful insights regarding the relationship between household debt and the Consumer Price Index (CPI) were sought by delving into the intricacies of the dataset. The analysis commenced with exploratory data analysis (EDA), whereby the distribution and interrelationships among variables were meticulously examined. Initial understanding of the data's patterns and associations was facilitated through visualizations such as scatter plots, histograms, and correlation matrices.

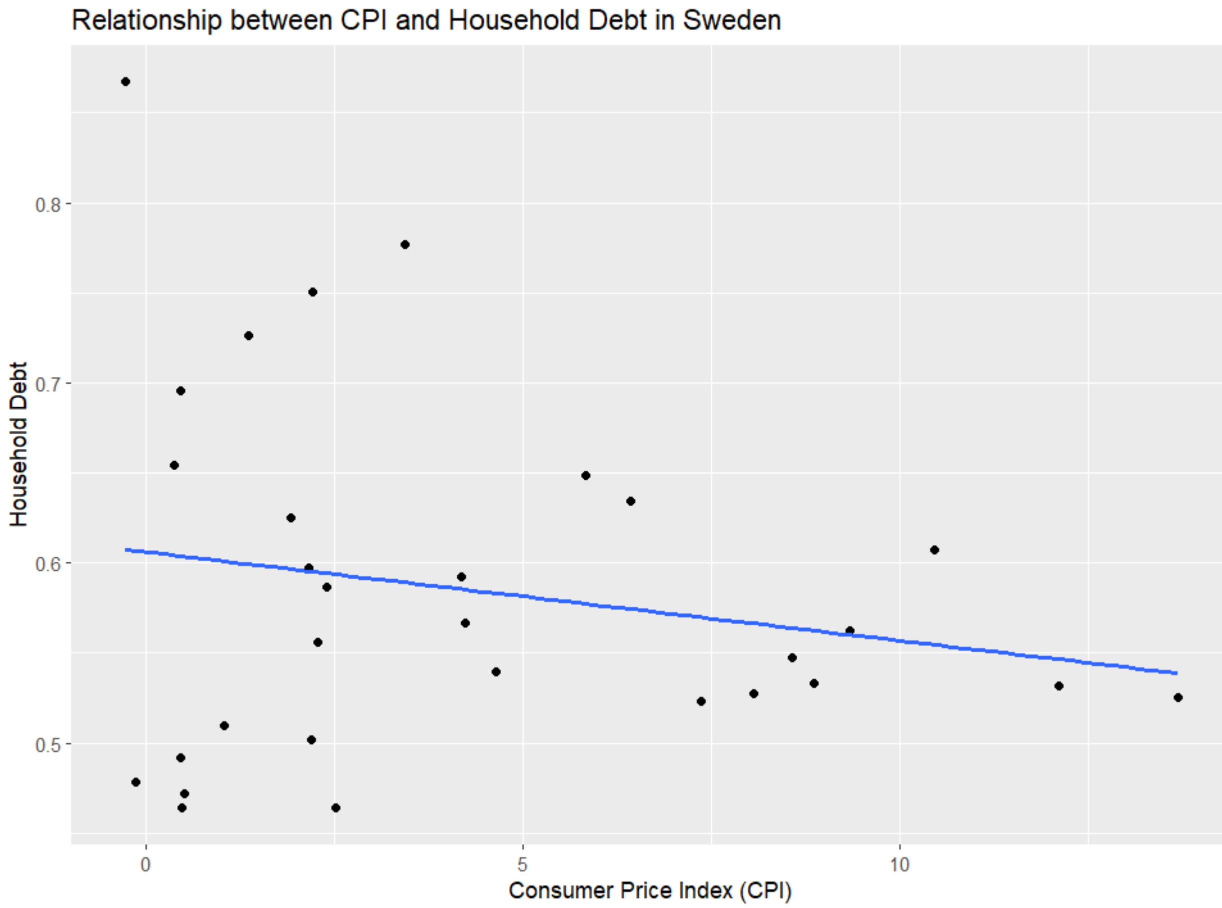

Figure 1,0: The relationship between household debt and the Consumer Price Index

The scatter plot illustrates the inverse relationship between the Consumer Price Index (CPI) and Household Debt in Sweden. As CPI increases, household debt tends to decrease. The negative slope of the trend line confirms this relationship.

```
lm(formula = debthhr ~ cpi, data = house_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.13950 -0.04600 -0.01410  0.05203  0.25997 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.606030   0.026882  22.544   <2e-16 ***
cpi         -0.004915   0.004687  -1.049    0.303    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09911 on 28 degrees of freedom
Multiple R-squared:  0.03779,	Adjusted R-squared:  0.003421 
F-statistic:   1.1 on 1 and 28 DF,  p-value: 0.3033
```
Figure 2.0 Analysis of Linear Regression Model for Household Debt and CPI

The coefficient for CPI suggests a negative association with household debt, although it is not statistically significant at the conventional significance level of 0.05. This implies that, on average, an increase in CPI does not reliably predict a decrease in household debt. Furthermore, the low R-squared value indicates that only a small proportion (approximately 3.78%) of the variance in household debt can be explained by changes in CPI. Overall, while there appears to be some relationship between CPI and household debt, other factors not accounted for in the model may have a more substantial influence on household debt dynamics. Further investigation into these additional factors is warranted to gain a comprehensive understanding of the determinants of household debt.

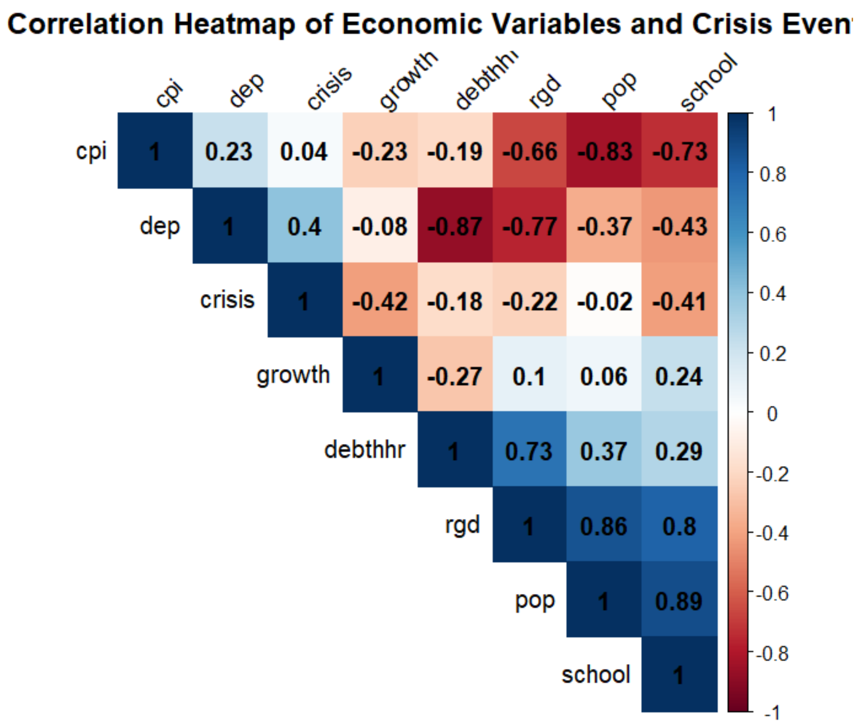

Figure 3: Correlation Among Economic Variables

A strong correlation is observed between household debt and Real GDP per capita, with a correlation coefficient of 0.73. Additionally, there is a notably high correlation of 0.87 with the dependency ratio. This suggests that factors beyond the Consumer Price Index (CPI) may significantly influence household debt levels. Consequently, when constructing an inference model for household debt, it is imperative to consider these variables alongside CPI to ensure a comprehensive understanding and accurate predictions.

Variable (Feature) Importance Analysis

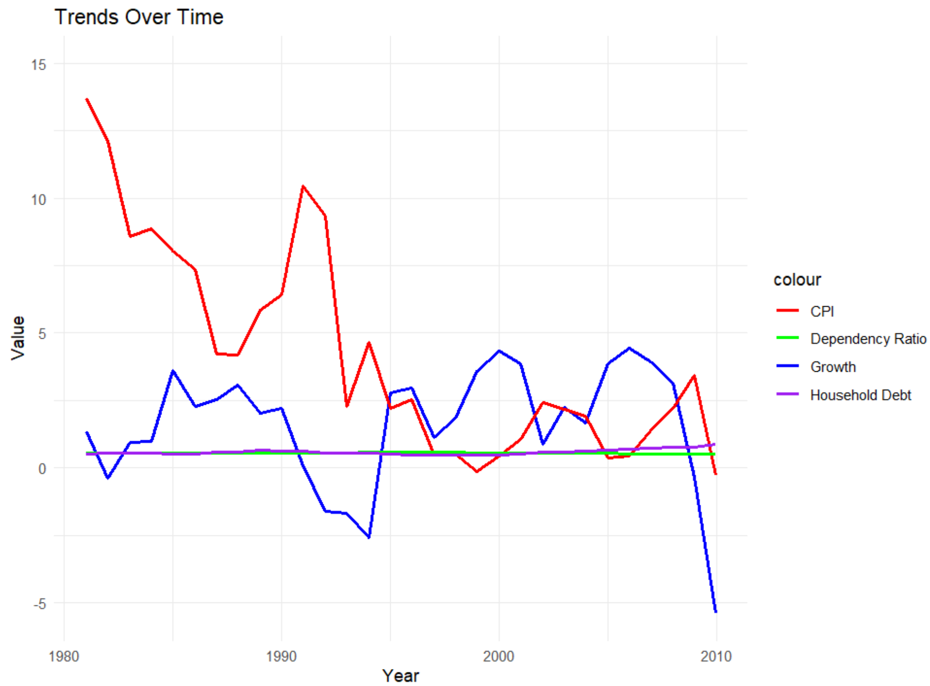

Figure 4.0: Trends in Economic Indicators Over Time with Crisis Periods Highlighted

Significant volatility is observed in the Consumer Price Index (CPI), with notable impacts on the economy. As the CPI rises, there is a negative effect on economic growth, while a decrease in the CPI corresponds to a decline in the economy.

Develop a Best Fit Model to Predict the house debt Data Using the Historical data

In this analysis, an attempt was made to develop a linear regression model to examine the relationship within the provided dataset concerning household debt. The selected parameters for constructing the model included the Consumer Price Index (CPI), Real GDP per capita, Real GDP per capita growth, Population, Average years of schooling, Dependency Ratio (dep), and Banking Crisis.

Multicollinearity

Despite achieving a high level of accuracy in the model, it was observed that multicollinearity exists among the variables. Figure 5.0 shows multicollinearity.

```
print(vif_values)
      cpi        rgd     growth        pop     school        dep     crisis
 5.087065  72.986312   1.544170 118.146084  31.707621  18.470550   6.939945
```

Figure 5.0 Multicollinearity with household debt and other variables.

The multicollinearity analysis highlights the correlations among predictors, emphasizing the importance of selecting variables with minimal collinearity for inclusion in the regression model. Notably, "rgd" (Real GDP per capita) stands out with a substantial coefficient of approximately 72.99, indicating a strong positive correlation with the response variable. Conversely, "cpi" (Consumer Price Index) shows a much lower coefficient of around 5.09, suggesting a weaker association.

Given the aim of reducing multicollinearity, variables such as "pop" (Population), "school" (Average Schooling Years), and "dep" (Dependency Ratio) demonstrate high coefficients, indicating significant correlations with other predictors. To address multicollinearity concerns, it's prudent to retain "rgd" while excluding highly correlated variables like "pop," "school," and "dep."

By prioritizing variables with less collinearity, we aim to enhance the robustness and interpretability of the regression model. This approach ensures clearer insights into the relationship between household debt and economic indicators while mitigating potential instability and ambiguity introduced by multicollinearity.

```
   cpi    rgd growth    dep crisis
2.8833 6.3505 1.3388 4.1322 1.5056
```

Figure 6.0 Multicollinearity with household debt without Population and the Average Schooling Years

Multicollinearity was addressed by regenerating the model without "pop" (Population) and "school" (Average Schooling Years), leading to improvement, as depicted in Figure 6.0.

Linear Regression Model with K-Fold Cross Validation

The 5-fold cross-validated linear regression model in Figure 7.0 demonstrates strong performance.

```
lm(formula = .outcome ~ ., data = dat)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.051172 -0.010903 -0.003549  0.017114  0.034204 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.318e+00  6.362e-01   3.644  0.00143 ** 
cpi         -1.859e-03  2.460e-03  -0.756  0.45773    
rgd          2.217e-05  7.901e-06   2.806  0.01029 *  
growth      -1.578e-02  2.349e-03  -6.716 9.47e-07 ***
pop         -4.439e-05  1.683e-04  -0.264  0.79438    
school      -6.800e-02  2.633e-02  -2.582  0.01700 *  
dep         -2.219e-02  1.293e-02  -1.716  0.10014    
crisis      -6.617e-02  3.263e-02  -2.028  0.05484 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02306 on 22 degrees of freedom
Multiple R-squared:  0.9591,	Adjusted R-squared:  0.946 
F-statistic: 73.64 on 7 and 22 DF,  p-value: 8.545e-14
```

```
Mean Squared Error (MSE): 0.000389970153496716
Cross-validated MSE: 0.000798574037571692
```

Figure 7.0 Linear Regression Model Summery values

With an adjusted R-squared of 0.946, it explains approximately 94.6% of household debt variability, indicating a robust fit. The model's low residual standard error (0.02306) suggests accurate predictions, while a small p-value (8.545e-14) confirms its statistical significance. These findings highlight the model's effectiveness in capturing household debt dynamics and underscore its reliability.

Lasso Regression Model with K Fold Cross Validation

```
Mean Squared Error (MSE): 0.0003917210314289960
Cross-validated MSE: 1.92727313177e-05
```

Figure 8.0: Lasso Regression Model Summery Analysis

The Lasso regression model demonstrates strong predictive accuracy, with low Mean Squared Error (MSE) values, indicating close alignment between predicted and actual household debt values. In summary, the model provides a concise yet accurate representation of household debt dynamics.

Redge Regression Model with K – Fold Cross Validation

The Ridge regression model shows strong predictive accuracy with a mean squared error (MSE) of approximately 0.000119. Cross-validation confirms its reliability, yielding a slightly higher MSE of about 0.000231, indicative of robust performance on unseen data. Overall, the Ridge model offers a reliable framework for understanding and predicting household debt dynamics.

```
Mean Squared Error (MSE): 0.0003917210314289960
Cross-validated MSE: 0.0276602177607074
```

Results:

The Linear Regression Model has the lowest MSE, meaning its predictions are closest to the actual values of household debt on average. However, its cross-validated MSE is higher, suggesting it might overfit to the training data. The Lasso Regression Model has a slightly higher MSE than the linear model but significantly lower cross-validated MSE. This indicates it performs well on unseen data, making it a strong choice. The Ridge Regression Model has the highest cross-validated MSE, suggesting it may not generalize well to new data despite its similar MSE to the Lasso model.

Considering these factors, the Lasso Regression Model seems to be the best choice. It balances prediction accuracy and generalization ability effectively, making it robust for predicting household debt. Table 1.0 describes the models Mean Squared Error and Cross Validation MSE.

Model	MSE	Cross-Validated MSE
Linear Regression	0.00038997	0.000798574
Lasso Regression	0.000391511	0.027787738
Ridge Regression	0.000391721	0.027660218

Table 1.0: Prediction Result on the Three Models

Prediction of the Households Debt data

Analysis for the year 2021 was conducted using available data. The Lasso and Ridge models yielded a prediction of 0.5851834 for household debt. Meanwhile, the Linear Regression model predicted a household debt value of 1.38 for the same year.