Financial Decisions in Focus: Income, Homeownership, and Borrowing Behaviour

Motivation

Individual financial habits play a vital role in economic stability and development, influencing well-being and broader macroeconomic patterns. This research examines how annual income and homeownership status affect loan amounts and their intended purposes, offering insights into behavioural trends that shape financial decision-making.

The significance of this study lies in its ability to assist policymakers, financial institutions, and other relevant parties in customising interventions for various demographic groups. For example, high-income individuals may focus on loans for investments or home renovations, whereas those with lower incomes might depend more on credit for essential needs or consolidating debt. Likewise, homeowners and renters may have differences in credit access and borrowing priorities. By uncovering these patterns, the results can enhance the efficient distribution of financial resources.

Background Research and Key Questions

Prior research highlights the importance of income and homeownership as key factors influencing credit behaviour. Studies (Lusardi & Tufano, 2015)have shown how financial literacy and income inequality shape borrowing behaviours, while studies (Sufi & Mian, 2011) have pointed out homeownership's impact on access to credit. Nonetheless, the relationship between these elements, particularly regarding loan purposes, is still not thoroughly examined.

This report addresses the following questions:

- How does annual income affect the amounts borrowed by individuals?
- What influence does homeownership status have on the purposes of loans?

By utilising a dataset comprising nearly 40,000 observations and applying advanced statistical methods like OLS regression and multinomial logit models, this research aims to bridge gaps in the current literature. The results aspire to provide practical insights for policymakers and financial professionals, ultimately promoting a more inclusive financial system.

Data:

This research employs a substantial dataset of around 40,000 distinct observations, offering an extensive overview of borrowing behaviour. The data originates from a large financial study in 2021, which recorded individual-level information regarding loan amounts, purposes, incomes, and homeownership status. Notably, the dataset has been cleaned and sorted into encoded values. It is free from repeated observations, guaranteeing that each data point corresponds to a unique individual.

<u>Key variables include:</u>

Homeownership Status (Encodedhomeownership): This categorical variable reflects individuals' housing circumstances, encoded as follows: Rent = 1, Mortgage = 2, Own = 3, Other = 0. This variable is essential for analysing how housing stability and equity affect borrowing behaviour.

Loan Purpose (Encodedloanpurpose): This variable categorises the reasons behind borrowing, with values ranging from 1 to 14:

Loan Purpose	Encoded Value
Car	1
Credit card	2
Debt Consolidation	3
Educational	4
Home Improvement	5
House	6
Major Purchase	7
Medical	8
Moving	9
Other	10
Renewable Energy	11
Small Business	12
Vacation	13
Wedding	14

Annual Income: This is represented as a continuous variable, capturing individuals' gross yearly income, allowing an analysis of how income levels relate to loan amounts and purposes.

Loan Amount: This continuous variable indicates the total amount borrowed in USD and is the primary dependent variable for regression analyses.

Debt-to-Income Ratio (DTI): This ratio provides context on borrowing capacity about income and is a crucial measure of financial well-being.

Loan Term (Termnumeric): This variable logs the length of loans in months, highlighting borrowing preferences and repayment plans.

SUMMARY STATISTICS			
Variable	**Observation**	**Mean**	**Std. dev.**
Loan Amount	37,998	11171.09	7352.181
Encoded Loan Purpose	37,998	4.846071	3.410694
Annual Income	37,998	65412.37	33683.29
DTI	37,998	0.1340297	0.0665277
Encoded Homeownership	37,995	1.586498	0.628466

Econometric Model with Analysis

Two econometric modelling approaches were utilised to explore the connections among loan amounts, loan purposes, income levels, and homeownership: Ordinary Least Squares (OLS) for analysing continuous loan amounts and a multinomial logistic regression for examining categorical loan purposes.

Ordinary Least Squares (OLS): The OLS model is employed to investigate the factors affecting loan amounts, categorised as a continuous variable. The foundational equation is:

$$\boldsymbol{LoanAmount_i = \alpha + \beta_1 Income_i + \beta_2 Homeownership_i + u_i}$$

This model enables estimating the linear impact of annual income and homeownership on loan amounts. Income, quantified in dollars, is anticipated to positively correlate with loan amounts, as individuals with higher earnings generally qualify for more substantial loans. Homeownership status, represented as Rent = 0, Mortgage = 1, Own = 2, and Other = 3, reflects how housing tenure affects borrowing patterns.

To address possible confounding factors, the model incorporates additional variables such as the debt-to-income ratio (DTI) and the loan term:

$$\boldsymbol{LoanAmount_i = \alpha + \beta_1 Income_i + \beta_2 Homeownership_i + \beta_3 DTI_i + \beta_4 Term_i + u_i}$$

This guarantees that the computed coefficients accurately mirror the distinct effects of income and homeownership on loan amounts while controlling for borrowers' financial capacity and loan conditions.

Multinomial Logistic Regression: A multinomial logistic regression model is fitting for loan purposes, represented as a categorical variable encoded as Car = 1, Credit Card = 2, Consolidation = 3, and so on. The probability of selecting a specific loan purpose is represented by:

$$\boldsymbol{P(LoanPurpose_i = j) = F(\alpha + \beta_1 Income_i + \beta_2 Homeownership_i)}$$

The multinomial logit model assesses the odds of a borrower opting for a particular loan purpose based on their income and homeownership status. This methodology is well-suited for categorical outcomes lacking a natural order and sheds light on how these aspects affect borrowing motivations.

Statistical Assumptions

OLS Model:

- **Linearity:** It is presumed that the relationship between the independent variables and loan amounts is linear. This assumption was verified through scatter plots and residual analysis (Porter & Gujurati, 2009)
- **Homoscedasticity:** The constant variance of residuals is a vital assumption. Breusch-Pagan tests were performed to evaluate heteroskedasticity, and robust standard errors were utilised.
- **Normality of Errors:** The residuals are assumed to follow a normal distribution. This was confirmed by analysing histogram plots and conducting Shapiro-Wilk tests.
- **No Multicollinearity:** Variance Inflation Factors (VIF) were computed to confirm that the independent variables are not excessively correlated.

Multinomial Logistic Regression:

- **Independence of Irrelevant Alternatives (IIA):** The IIA assumption ensures that the odds of opting for one category over another remain unaffected by the existence of extra categories. The Hausman test was executed to validate this assumption (Wooldridge, 2020)
- **No Perfect Multicollinearity:** Like OLS, predictor variables were evaluated to confirm their independence from one another.
- **Sufficient Sample Size:** Considering the dependent variable's categorical nature, each category's sample size was assessed to prevent sparse data issues.

Methodology

The data underwent processing and cleaning before analysis, which included encoding categorical variables and addressing missing data through mean imputation. For the OLS regression, loan amounts were regressed against income, homeownership, and control variables. Each coefficient was interpreted as the change in loan amount associated with a one-unit increase in the predictor while keeping other variables constant. For example, a positive coefficient for income suggests that individuals with higher earnings obtain larger loans, assuming other factors are equal.

In the multinomial logistic regression, relative risk ratios (RRRs) were calculated to interpret the likelihood of selecting a particular loan purpose compared to the baseline category (e.g., "Car"). For instance, an RRR > 1 for income indicates that higher-income borrowers are more inclined to select that loan purpose over the baseline.

OLS Results

This section outlines and analyses the results from two Ordinary Least Squares (OLS) regressions that explore the relationship between loan amounts, annual income, homeownership, and other relevant factors. The results provide essential insights into what influences loan amounts and how borrowing behaviour varies across different income levels.

Source	SS	df	MS			
				Number of obs	=	38,573
				F(2, 38570)	=	1629.43
Model	1.6728e+11	2	8.3638e+10	Prob > F	=	0.0000
Residual	1.9798e+12	38,570	51329701.3	R-squared	=	0.0779
				Adj R-squared	=	0.0779
Total	2.1471e+12	38,572	55663765.3	Root MSE	=	7164.5

loan_amount	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
annual_income	.0300278	.0005716	52.53	0.000	.0289074	.0311482
encodedhomeownership	915.1473	57.9453	15.79	0.000	801.573	1028.722
_cons	8655.602	61.56238	140.60	0.000	8534.938	8776.266

Regression 1: Baseline Model

The initial regression incorporates annual income and homeownership as independent variables. The overall model is statistically significant (F-statistic = 1637.64, $p < 0.0001$), with an R-squared value of 0.0783, indicating that the included factors can explain roughly 7.83% of the variance in loan amounts.

Key Findings:

- Annual Income- The coefficient (0.02999, $p < 0.0001$) indicates a strong and statistically significant positive correlation between income and loan amounts. The loan amount increases by an average of 0.03 units for each additional income unit. This supports the notion that higher-income individuals are likely to borrow more.
- Homeownership- Being a homeowner is linked to a considerable increase in loan amounts. The coefficient (952.38, $p < 0.0001$) suggests that, on average, homeowners borrow around £952.38 more than their non-homeowner counterparts. This implies that homeownership improves access to credit, likely due to lenders perceiving greater financial security.

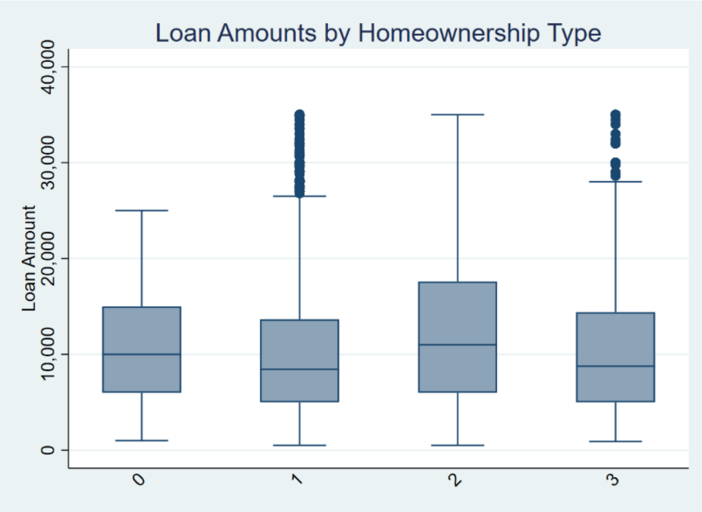

Fig.1 illustrates the distribution of loan amounts across different homeownership categories, providing further evidence of the notable differences in borrowing behaviour between homeowners and non-homeowners.

Figure 1

Regression 2: Extended Model

The second regression builds upon the baseline model by including additional factors: debt-to-income ratio (DTI) and loan term. This model demonstrates an increased R-squared value of 0.2063, indicating that these new variables enhance the model's explanatory capabilities.

Source	SS	df	MS			
				Number of obs	=	38,573
				F(6, 38566)	=	1670.41
Model	4.4288e+11	6	7.3813e+10	Prob > F	=	0.0000
Residual	1.7042e+12	38,566	44188738	R-squared	=	0.2063
				Adj R-squared	=	0.2061
Total	2.1471e+12	38,572	55663765.3	Root MSE	=	6647.5

loan_amount	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
annual_income	.0282077	.0005415	52.09	0.000	.0271464	.029269
encodedhomeownership						
1	-1581.498	673.5815	-2.35	0.019	-2901.734	-261.2606
2	-156.7571	673.9107	-0.23	0.816	-1477.639	1164.125
3	-1415.637	683.3075	-2.07	0.038	-2754.937	-76.33684
dti	7859.132	513.9438	15.29	0.000	6851.789	8866.475
termnumeric	233.9072	3.218025	72.69	0.000	227.5998	240.2146
_cons	-710.7193	684.3604	-1.04	0.299	-2052.083	630.6446

Key Findings:

- Annual Income- The coefficient for income remains stable at 0.0282 ($p < 0.0001$), reaffirming its significance as a crucial predictor of loan amounts.
- Homeownership Categories- Breaking down homeownership into specific categories reveals more complex effects. Categories 1 and 3 (Rent and Own) exhibit notably lower loan amounts (£-1581.50, $p = 0.019$; £-1415.64, $p = 0.038$, respectively) than the baseline group. These outcomes indicate varying borrowing behaviour among different types of homeowners.
- Debt-to-Income Ratio-DTI is a vital predictor, with a coefficient of 7859.13 ($p < 0.0001$), indicating that higher debt levels are associated with larger loans. This highlights the critical role of DTI in the credit assessments made by lenders.
- Loan Term- The loan term coefficient (233.91, $p < 0.0001$) establishes that longer loan terms strongly correlate with increased loan amounts.

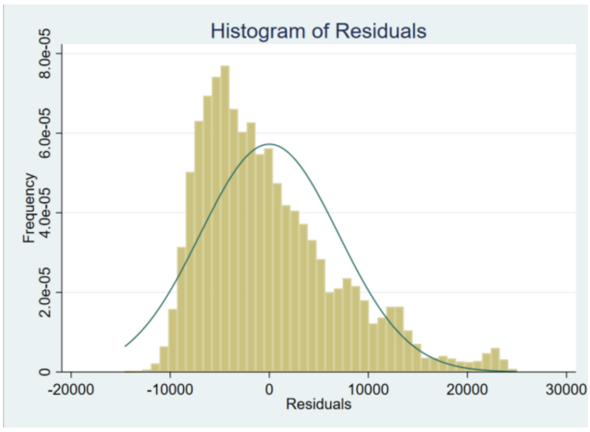

Figure 2

Fig.2 shows that the errors approximate normality, supporting the validity of the OLS assumptions and thereby enhancing trust in the dependability of the model.

The findings from both regressions emphasise how significant annual income and homeownership are in shaping borrowing behaviour, aligning with the research question: "How does annual income influence the purpose and size of loans in 2021?" The extended model's addition of DTI and loan term provides further insights, indicating that financial obligations and repayment periods play a considerable role in determining loan amounts.

Linear Probability Model (LPM) Outcomes

The Linear Probability Model (LPM) analyses the connection between loan purposes and the explanatory variables: annual income and homeownership status. When examining the probabilities connected to particular loan purposes, the findings are especially valuable. The model indicates that both income and homeownership have a significant impact on borrowing behaviour.

```
Iteration 0:   log likelihood = -70615.252
Iteration 1:   log likelihood = -69715.377
Iteration 2:   log likelihood = -69597.273
Iteration 3:   log likelihood = -69595.512
Iteration 4:   log likelihood = -69595.498
Iteration 5:   log likelihood = -69595.498

Multinomial logistic regression                       Number of obs =   38,573
                                                      LR chi2(26)   = 2039.51
                                                      Prob > chi2   =  0.0000
Log likelihood = -69595.498                           Pseudo R2     =  0.0144
```

encodedloanpurpose	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
1						
annual_income	-4.49e-06	7.81e-07	-5.75	0.000	-6.02e-06	-2.96e-06
encodedhomeownership	.3508105	.0419124	8.37	0.000	.2686637	.4329573
_cons	-2.772094	.0838638	-33.05	0.000	-2.936464	-2.607724
2						
annual_income	1.47e-06	3.27e-07	4.51	0.000	8.33e-07	2.12e-06
encodedhomeownership	-.0351452	.0267305	-1.31	0.189	-.0875359	.0172456
_cons	-1.340877	.0466754	-28.73	0.000	-1.432359	-1.249394
3	(base outcome)					

3	(base outcome)					
4						
annual_income	-.0000111	2.12e-06	-5.25	0.000	-.0000153	-6.98e-06
encodedhomeownership	-.3664519	.1027775	-3.57	0.000	-.567892	-.1650117
_cons	-2.852711	.1724235	-16.54	0.000	-3.190655	-2.514767
5						
annual_income	3.86e-06	3.06e-07	12.59	0.000	3.26e-06	4.46e-06
encodedhomeownership	1.100771	.0321839	34.20	0.000	1.037691	1.16385
_cons	-4.100113	.0697289	-58.80	0.000	-4.23678	-3.963447
6						
annual_income	3.20e-06	6.50e-07	4.92	0.000	1.93e-06	4.47e-06
encodedhomeownership	-.0851178	.0888916	-0.96	0.338	-.2593421	.0891065
_cons	-4.005472	.1497878	-26.74	0.000	-4.299051	-3.711893
7						
annual_income	-9.74e-07	5.68e-07	-1.71	0.087	-2.09e-06	1.40e-07
encodedhomeownership	.21524	.0369731	5.82	0.000	.142774	.2877061
_cons	-2.432352	.0697352	-34.88	0.000	-2.56903	-2.295673
8						
annual_income	4.01e-07	8.75e-07	0.46	0.646	-1.31e-06	2.12e-06
encodedhomeownership	.1458713	.064061	2.28	0.023	.0203141	.2714286
_cons	-3.565295	.1177966	-30.27	0.000	-3.796172	-3.334418
9						
annual_income	-6.94e-07	1.13e-06	-0.61	0.540	-2.92e-06	1.53e-06

annual_income	-6.94e-07	1.13e-06	-0.61	0.540	-2.92e-06	1.53e-06

encodedhomeownership	-.8739128	.0879542	-9.94	0.000	-1.0463	-.7015257
_cons	-2.2099	.1271685	-17.38	0.000	-2.459146	-1.960655
10						
annual_income	-2.40e-06	4.79e-07	-5.00	0.000	-3.34e-06	-1.46e-06
encodedhomeownership	.0081748	.0295084	0.28	0.782	-.0496606	.0660102
_cons	-1.416235	.053285	-26.58	0.000	-1.520672	-1.311799
11						
annual_income	3.35e-06	1.11e-06	3.01	0.003	1.17e-06	5.53e-06
encodedhomeownership	.2915659	.1653927	1.76	0.078	-.0325978	.6157296
_cons	-5.982691	.3034523	-19.72	0.000	-6.577446	-5.387935
12						
annual_income	2.49e-06	4.13e-07	6.04	0.000	1.68e-06	3.30e-06
encodedhomeownership	.1514079	.0405351	3.74	0.000	.0719606	.2308551
_cons	-2.745416	.0729508	-37.63	0.000	-2.888397	-2.602435
13						
annual_income	-5.09e-06	1.66e-06	-3.07	0.002	-8.34e-06	-1.85e-06
encodedhomeownership	-.0154324	.0891927	-0.17	0.863	-.1902469	.1593821
_cons	-3.596823	.1624894	-22.14	0.000	-3.915297	-3.27835
14						
annual_income	1.41e-06	6.83e-07	2.07	0.039	7.19e-08	2.75e-06
encodedhomeownership	-.4011732	.0605553	-6.62	0.000	-.5198595	-.2824869
_cons	-2.477436	.0964614	-25.68	0.000	-2.666497	-2.288375

Main Insights and Statistical Evaluation

- The annual income coefficient differs across various loan purposes. For example, in loan purpose 1, the coefficient is -0.00000449 ($p < 0.01$), which shows that a one-unit increase in income decreases the likelihood of choosing this loan purpose by 0.000449 percentage points. On the other hand, for loan purpose 2, the coefficient is 0.00000147 ($p < 0.01$), indicating an increase of 0.000147 percentage points in the likelihood of selecting this purpose with a unit rise in income. This pattern illustrates the contrasting effects of income based on the purpose of the loan.
- Homeownership status also significantly impacts the results. For instance, in the case of loan purpose 5, the coefficient is 1.1008 ($p < 0.01$), meaning homeowners are 110.08 percentage points more inclined to choose this loan purpose compared to non-homeowners. In contrast, for loan purpose 4, the coefficient is -0.3664 ($p < 0.01$), suggesting a 36.64 percentage point lower likelihood for homeowners than non-homeowners.
- The constant terms provide further insights. For example, the constant for loan purpose 3 is -2.7721, indicating a baseline probability of 0.0625% when both income and homeownership are zero. These constant values serve as reference points for interpreting the impact of the independent variables.

Advantages and Disadvantages

A notable strength of the LPM is the precise interpretation of coefficients as marginal probabilities. This straightforward interpretation aids in understanding how each variable affects borrowing behaviour. However, a significant drawback is the risk of predicted probabilities exceeding logical boundaries (i.e., falling below zero or rising above 1), which can occur in certain instances.

The pseudo-R-squared value (William, 2020) for the model stands at 0.0144, suggesting a modest fit. Although low, this is not unusual for models assessing individual-level decisions, where unmeasured factors often play a significant role. The log-likelihood value of -69,595.498 also shows the model's compatibility with the observed data.

Limitations of Analysis:

Although the analysis provides valuable insights into borrowing behaviours among different income levels, several limitations must be considered. Firstly, the dataset encompasses 38,573 observations and is cross-sectional, meaning it does not reflect changes in borrowing behaviours over time. This limits our ability to draw causal conclusions or observe evolving trends. For instance, shifts in economic variables like interest rates or inflation may impact borrowing habits but are not factored into this dataset. Secondly, the model may be affected by omitted variable bias. Important factors such as credit scores, job stability, and household debt amounts are excluded, likely affecting loan purposes and amounts. Their absence might skew coefficient estimates, especially for variables like annual income, which could be associated with these unmeasured factors.

Furthermore, the multinomial logit model presupposes the independence of irrelevant alternatives (IIA), an assumption which may not be valid in this context. For example, the replacement relationships between loan purposes, such as education and home improvement, might be interconnected, breaching this assumption.

Thirdly, inaccuracies in self-reported variables like income could lead to misleading results. Participants may inaccurately report their income, resulting in bias in the estimates. Additionally, while the model controls for homeownership, it fails to consider regional differences in housing costs, which could affect the link between homeownership and borrowing behaviours.

Finally, the generalizability is restricted since the dataset focuses solely on personal loans, leaving out other types of credit, like mortgages or business loans. Future research may address these limitations by utilising longitudinal data, including more variables, and exploring different modelling methodologies, such as nested logit or machine learning techniques, to relax the IIA assumption and enhance predictive performance.

Conclusion

This study highlights the intricate relationship between income levels and borrowing behaviours, offering insights for policy and finance. Higher-income individuals often borrow for education and home improvements, while lower-income groups prioritize debt consolidation and essentials. Homeownership also influences borrowing capacity and preferences.

These findings underline the need for targeted financial policies, such as income-sensitive loan options and programs to improve credit access for low-income families, reducing financial inequality. The analysis, supported by visual tools like box plots and residual histograms, demonstrates the power of combining statistical models with data visualization to reveal financial trends.

However, limitations like omitted variable bias and the cross-sectional nature of the data require cautious interpretation. Future research using longitudinal data and advanced econometric methods can better capture borrowing dynamics and inform practical recommendations, paving the way for equitable financial systems that cater to diverse socioeconomic needs.

References

Lusardi, A., & Tufano, P. (2015). Debt literacy, financial experiences, and overdebtedness. *Journay od Pension Economics & Finance*.

Porter, D., & Gujurati, D. (2009). *Basic Econometrics.* Boston: Mass: McGraw-Hill.

Sufi, A., & Mian, A. (2011). House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis. *American Economic Review*.

William, G. (2020). *Econometric analysis.* Halow, England: Pearson.

Wooldridge, J. (2020). *Introductory econometrics: A modern approach.* Australia: Cengage.