ETHIOPIA: AGRICULTURE & CREDIT ACCESS

ABSTRACT
In markets where credit approval rates are high, the loan principal may be the most meaningful variable for firms and households. This report aims to detail which characteristics of households contribute to the amount loaned to those approved for credit.

Majé Kendall
Econometrics ECN 480

Contents

Introduction

Access to capital is crucial to the growth and development of any society. This is especially true for emerging markets. This document will explore and report the factors which contribute most to the amount loaned to citizens in Ethiopia. Our report will be based on The Ethiopian Socioeconomic Survey conducted in 2013.

The Ethiopian Socioeconomic Survey (ESS) data was collected in 2013–14 from a nationally representative sample of households. Households were asked about topics such as their housing conditions, assets, and access to credit. In this report well explore the characteristics of households and then detail the characteristics of those approved for a loan in the last 12 months.

The analysis will end with a multiple regression analysis of variables found to be significant contributors to our variable in question, the amount loaned. The reader will be guided through the methodology used to construct the final model.

In the end, I will justify decisions made by validating the model, verifying assumptions, and offering final thoughts and conclusions.

Research Question

What factors contribute to the probability of approval for loan applicants? Is this an important question considering the approval rating? Is "what factors contribute to approval amount?" a better question?

This dataset offered a lot of flexibility in arriving towards a final research question. In fact, the question changed after the discovery of certain characteristics of our data. Initially, the question was, "Which factors contribute to the probability of being approved for a loan?". But upon further exploration found that there may be a more meaningful question:

<div align="center">"Which factors contribute to loan amounts and how much?"</div>

Data

The data in this report will be based on The Ethiopian Socioeconomic Survey conducted in 2013. The Ethiopian Socioeconomic Survey (ESS) is a collaborative project between the Central Statistics Agency of Ethiopia (CSA) and The World Bank. The objective of the study was to collect multi-topic panel household-level data with a special focus on improving agriculture statistics and the link between agriculture and other sectors of the economy. This report is based on one iteration of the data collected in 2013.

The Ethiopia Socioeconomic Survey 2013/2014 covered all regional states including the capital, Addis Ababa and is nationally representative. However, the sample is not representative of each of the small Ethiopian regions including Afar, Benshangul Gumuz, Dire Dawa, Gambella, Harari, and Somali regions.

The World Bank notes that estimates can be produced for a combination of all smaller regions as one "other region" category.

Our analysis was conducted using two tables. One a table containing all households, their characteristics, and whether they applied for a loan or not. Two, a table containing only the households which were approved for a loan and the characteristics of their loan and household. We made conversions to the loan amount from the Ethiopian Birr to USD. The exchange rate used was for the date Jan 1, 2015: 1 USD to .05 ETB. See appendix A for column descriptions and links to the data file.

Exploratory Data Analysis

I began with exploratory data analysis by creating summary and descriptive statistics for the variables. I separated the exploration into two interrelated parts—The household and the loan. The household section provides descriptive statistics for characteristics relating to the household in and of itself. There are various groups created on aggregate functions that give a more detailed look at how certain variables differ by categorical group. Scatter plots related to loans were used in simple regression so were moved to The Model section.

The Household

Here we will provide a descriptive overview of variables relating to the household. We found that age had a mean of 44.15 with a standard deviation of 15.60. We also calculated a mean by region. The mean household size was 4.57 with a SD of 2.40. The mean maximum education in each household was 7.54 with a SD of 7.27. The mean number of assets is 14.95 with a SD of 17.29. Most importantly, we saw that when we compare the characteristics of households who were approved for loans to those who were denied, the only variable with a significant difference between them is age.

Mean age by region

Addis Ababa	Afar	Amhara	Benshagul Gumuz	Diredwa	Gambelia
45.25170	41.52206	45.51917	41.72000	42.68182	42.08462
Harari	Oromia	SNNP	Somalie	Tigray	
44.90741	43.41555	43.04366	46.32753	45.58649	

Mean household size by region

Addis Ababa	Afar	Amhara	Benshagul Gumuz	Diredwa	Gambelia
3.976271	5.080882	4.080471	4.864000	4.109091	5.169231
Harari	Oromia	SNNP	Somalie	Tigray	
4.722222	4.712368	4.952991	5.156250	4.320132	

Mean education by region

Addis Ababa	Afar	Amhara	Benshagul Gumuz	Diredwa	Gambelia
14.291525	5.147059	6.583415	6.296000	7.895455	8.250000
Harari	Oromia	SNNP	Somalie	Tigray	
7.864198	7.999521	6.958582	4.118056	8.260297	

Mean number of Assets owned by region

Addis Ababa	Afar	Amhara	Benshagul Gumuz	Diredwa	Gambelia
23.325424	9.507353	18.049068	12.656000	12.172727	11.092308
Harari	Oromia	SNNP	Somalie	Tigray	
14.172840	12.613615	10.488471	7.809028	25.457990	

Differences in Household Characteristics—Confidence intervals
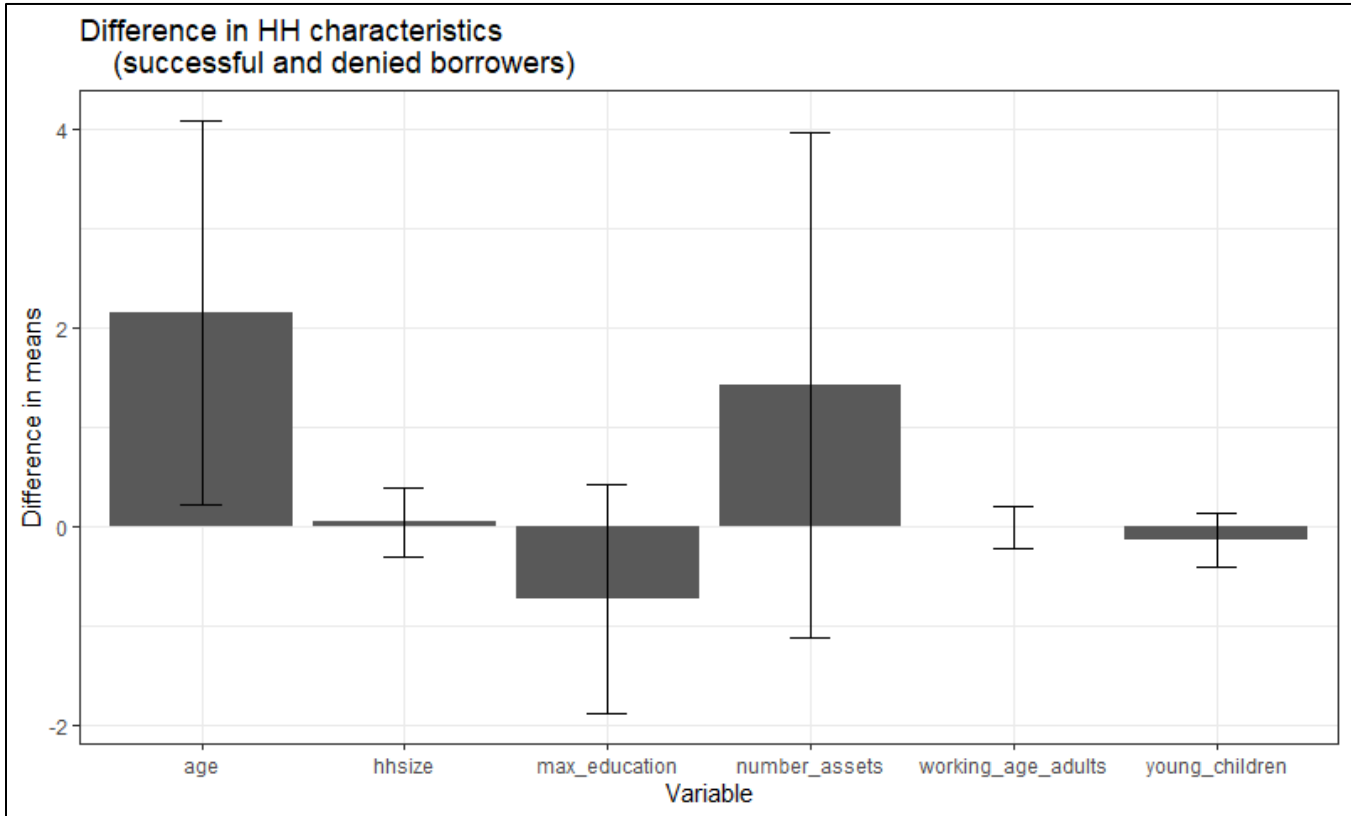
The Loan

Here we will provide a descriptive overview of variables relating to loan applications and agreements. The overall approval rate for loans was 87.14%. The approval rate by region shows a similar distribution. The histogram shows the frequency distribution of our loans in $50 increments. We see from our quartile analysis that the mean loaned amount is $163. However, there is a high degree of skewness in the data so we will rely on the median as a central point of tendency which is $60. Also, 75% of loans are less than $165.

Approval rate of sample

denied	successful
0.1285166	0.8714834

Approval rate by region

```
                      denied  successful
Addis Ababa           0.07407407 0.92592593
Afar                  0.11111111 0.88888889
Amhara                0.10614525 0.89385475
Benshagul Gumuz       0.18181818 0.81818182
Diredwa               0.12280702 0.87719298
Gambelia              0.21951220 0.78048780
Harari                0.10526316 0.89473684
Oromia                0.15488215 0.84511785
SNNP                  0.14905149 0.85094851
Somalie               0.07894737 0.92105263
Tigray                0.10400000 0.89600000
```

Loan Amount
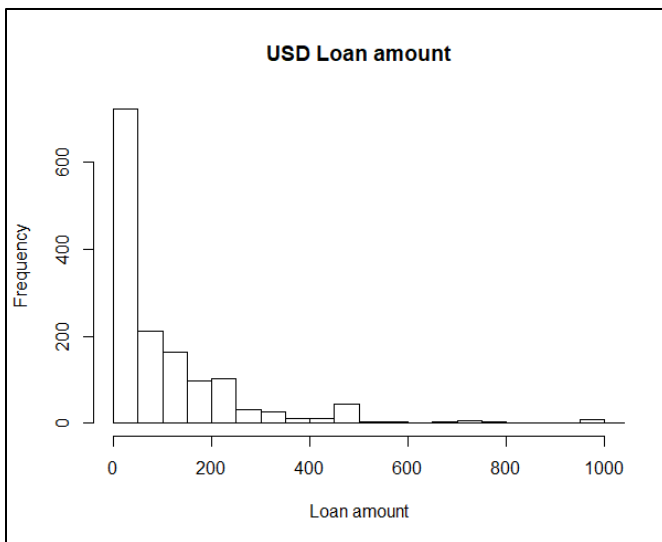

```
summary(gotL2$loan_amount_usd)
 Min. 1st Qu.  Median    Mean 3rd Qu.     Max.
 0.05   20.00   60.00  163.83  165.00  7500.00
```

The Model

In this section well walk through the methodology in our regression analysis. Our final model can be defined below:

$$LoanAmount = \beta_0 + \beta_1 age + \beta_2 education + \beta_3 numassets + \varepsilon$$

age is age in years
education is number of years of schooling
numassets is number of assets owned by household
ε is unobserved factors

We progressed through the analysis by first identifying variables that showed a relationship with the dependent variable or exhibited some level of significance when regressed against our dependent variable in a simple model. We chose a simple functional form because there didn't seem to be a non-linear pattern associated with any of our independent variables. See appendix B for our simple regression models and their respective outputs.

Multiple Regression Results

```
Coefficients:
              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
(Intercept)     79.783    32.3447   2.467 1.375e-02   16.336 143.2292 1469
number_assets    4.410     1.1240   3.923 9.138e-05    2.205   6.6145 1469
max_education    8.574     2.4587   3.487 5.029e-04    3.751  13.3967 1469
age             -1.107     0.6332  -1.748 8.074e-02   -2.349   0.1355 1469

Multiple R-squared:  0.06816 ,  Adjusted R-squared:  0.06625
F-statistic: 12.14 on 3 and 1469 DF,   p-value: 7.581e-08
```

After selecting our variables, we immediately see that the model is significant with a p-value of .00000007581. We can be confident that we don't have a type I error. As we dive deeper into the output, we find that age cannot be considered to affect loan amounts. This mirrors the result of our simple model as well. However, our other variables tell a different story. Max education and number of assets in the household are variables that have been found to contribute to loan amount. Lastly, the model's R^2 of 0.06 shows that the variance of the dependent variable is not well explained by the dependent variables in the model. Because we are more interested in the model's causal effects instead of its predictive power, this doesn't hinder our results.

Assumptions

For this model, we considered the standard Gauss-Markov assumptions for multiple linear regression which are listed below:

- Linear in Parameters
- Random Sampling
- No Perfect Collinearity
- ε Zero Conditional Mean

Conclusion

I began my journey through this data set expecting to pursue one question, "Which factors contribute to loan approval rates?" But after instigating I quickly found that a more meaningful question my need to be answered, "Which factors contribute to loan amounts and how much?".

After analysis, we found two-thirds of the variables added to the regression model make a non-trivial contribution to loan amount. With *education* and *numassets* having respective coefficients of 8.57 and 4.41, these two variables can make a meaningful difference considering that 75% of loans are less than $165.

These results spark further discussion into why these factors contribute to loan amount. Are borrowers putting assets up for collateral? Can that make a difference in emerging markets that are thirsty for capital? Although the data set limits the number of questions we can answer, we could always seek data from other sources to continue the search for answers.

Appendix

Appendix A: Survey Data

Socioeconomic Survey 2013-2014 : https://microdata.worldbank.org/index.php/catalog/2247/study-description

Metadata and column descriptions:

Variable name	Description
household_id2	Unique household ID
got_loan	Over the past 12 months, did anyone in this household borrow on credit from someone outside the household for business or farming purposes?
rural	Whether household lives in rural area, small town (urban), or large town (urban)
hhsize	Household size
region	Region
gender	Gender of household head
age	Age of household head (years)
young_children	Number of young children in household (13 years of age or below)
working_age_adults	Number of working-age adults in household (aged between 14 to 54)
max_education	Highest education level attained by anyone in the household (number of years of schooling)
number_assets	Number of assets owned by household (the sum of all items in Section 10 of the questionnaire)
Variables for 'All households' only	
loan_rejected	During the last 12 months, did anyone in the household try to borrow from someone outside the household and were turned down?
rejection_source1	Whom the household tried to get the loan from (source #1)
rejection_source2	Whom the household tried to get the loan from (source #2)
loan_purpose	Main purpose of loan
loan_purpose_other	Main purpose of loan (Other: Please specify)
did_not_apply	Did not apply for a loan within the last 12 months
reason_not_apply1	Reason for not applying (most important)
reason_not_apply2	Reason for not applying (second most important)
Variables for 'Got loan' only	
borrowed_from	Whom the household borrowed from
borrowed_from_other	Whom the household borrowed from (Other: Please specify)
loan_purpose	Main purpose of loan
loan_startmonth	When the loan was obtained (month)
loan_startyear	When the loan was obtained (year)
loan_repaid	Whether the loan has been repaid (at time of survey) or not
loan_endmonth	When (approximately) the loan is expected to be repaid (month)
loan_endyear	When (approximately) the loan is expected to be repaid (year)
loan_amount	Loan amount (principal), in birr (national currency)
loan_interest	Interest charged on loan, in birr

Appendix B: Simple Regression Models

Simple Regression: Loan Amount on Age
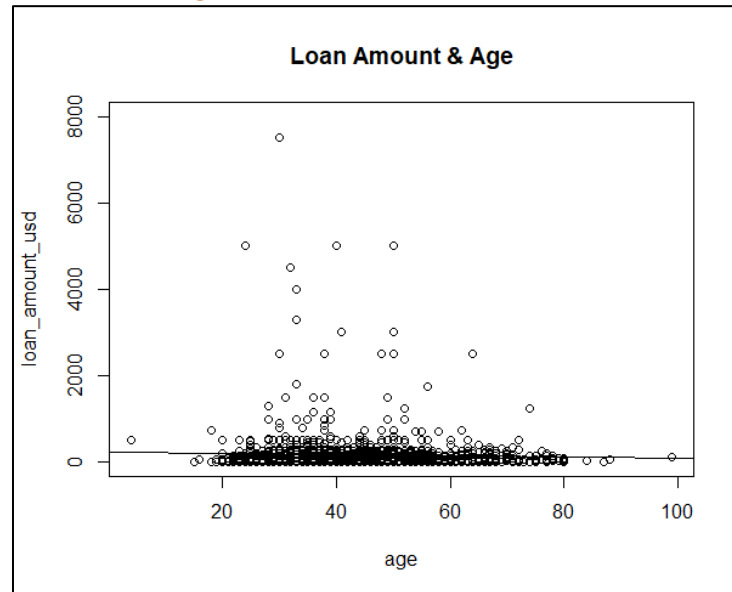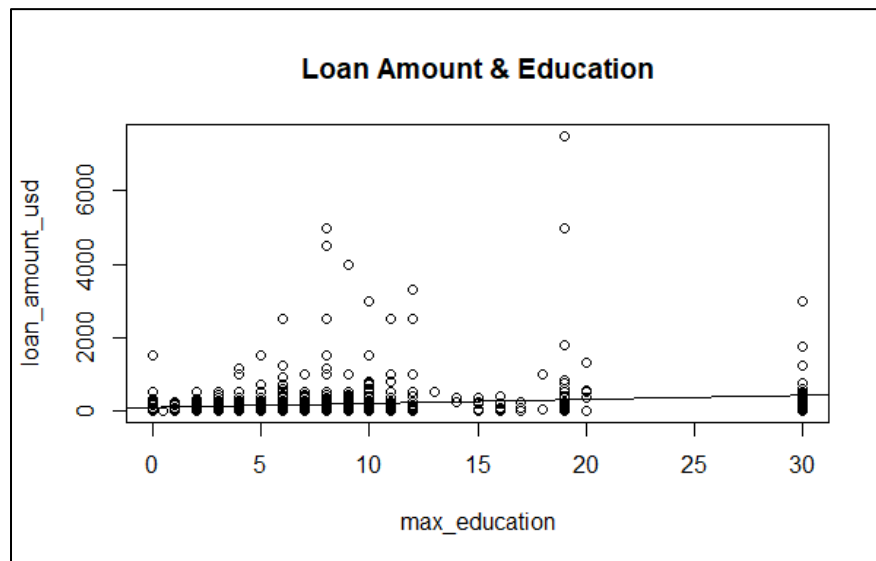


```
Coefficients:
            Estimate Std. Error t value  Pr(>|t|) CI Lower  CI Upper   DF
(Intercept)  219.39    36.4661    6.016 2.250e-09  147.855 290.91738 1471
age           -1.28     0.6864   -1.865 6.238e-02   -2.626   0.06627 1471

Multiple R-squared:  0.001738 , Adjusted R-squared:  0.00106
F-statistic: 3.478 on 1 and 1471 DF,  p-value: 0.06238
```

Simple Regression: Loan Amount on Education


```
Coefficients:
              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
(Intercept)     82.08    12.284    6.682 3.339e-11   57.983   106.17 1472
max_education   11.64     2.496    4.666 3.350e-06    6.749    16.54 1472

Multiple R-squared:  0.03188 ,  Adjusted R-squared:  0.03122
F-statistic: 21.77 on 1 and 1472 DF,  p-value: 3.35e-06
```
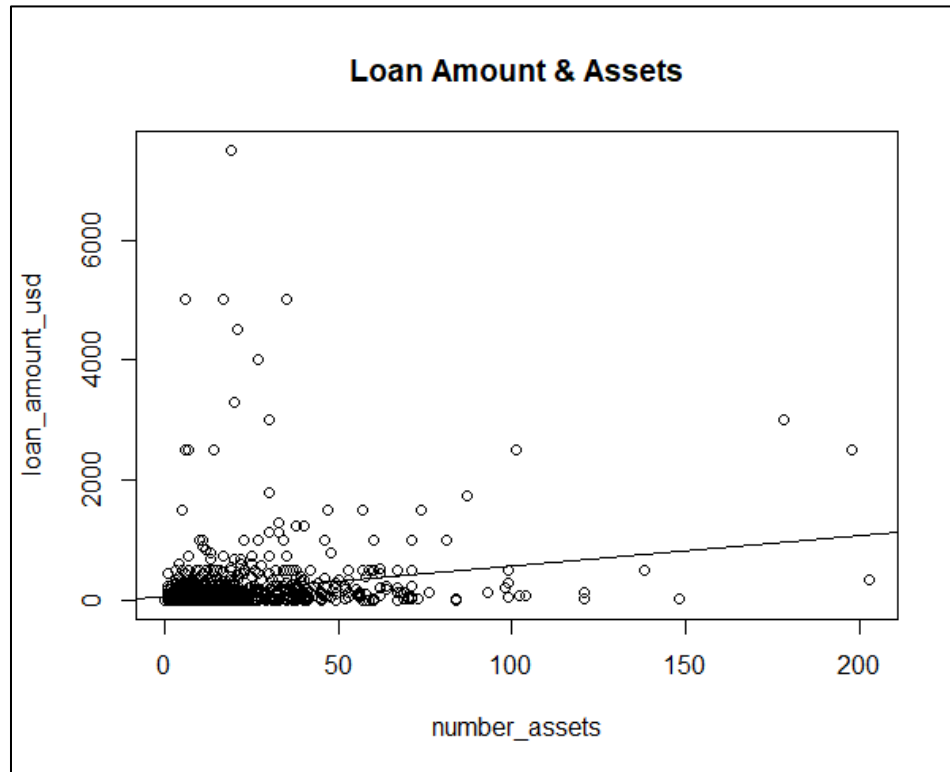
Simple Regression: Loan Amount on Number of Assets


```
Coefficients:
              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
(Intercept)     82.188     17.152   4.792 1.819e-06   48.544  115.832 1472
number_assets    5.012      1.084   4.624 4.090e-06    2.886    7.138 1472

Multiple R-squared:  0.04914 ,  Adjusted R-squared:  0.04849
F-statistic: 21.38 on 1 and 1472 DF,  p-value: 4.09e-06
```