Examining the Influence of Microfinance on Household Conditions in Hyderabad, India

Tara Fleming*

ABSTRACT

While predicting whether or not a person will default on a loan is nigh unto impossible, modeling the probability that they will incur back-debt by the frequency of missing payments could be a reasonable proxy for this measure. This paper aims to do this by regressing the number of loans that have gone into default during any cycle on a host of measures designed to assess the individual's ability to pay back the loans. The observations have been drawn from a survey of 2,800 households in India. I am interested to see if my logical predictions of which characteristics will cause an individual to fall into default will be proven accurate. The results of this analysis could potentially help microfinance institutions make more-informed decisions regarding whom to provide loans to, in order to minimize their risk factors. Moreover, I am intrigued to analyze the theory that females make better loan candidates than males. On one hand, it does seem logical that their business revenues will be reinvested within the family, engendering more consumption in the community. But will the increase in disposable income from successful businesses affect both genders of borrowers the same?

INTRODUCTION

The aim of microfinance is to make simple financial services available to impoverished peoples who otherwise would not have access to it. Micro-loans, -savings, and –insurance are all types of microfinance offered to people living in poverty. The loans are typically small and their maturity lengths are shorter than average, usually about one year. Microloan stipulations tend to be quite strict, requiring their recipients to attend weekly meetings for installment paybacks and business or personal finance training. Across many sources, it is found that around seventy percent of those living in the most extreme situations of poverty are women. Many microfinance institutions target their loans to these women because they are believed to be more likely to invest their loaned funds in conservative business ventures and to reinvest profits in their families or children's educations, thus increasing the social wellbeing of the entire community. Does this make their ability to repay loans greater than the men? Beyond logicality, does favoritism towards women reward institutions with profits? How can one tell?[*] While predicting whether or not a person will default on a loan is nigh unto impossible, modeling the probability that they will incur back-debt by the frequency of missing payments could be a reasonable proxy for this measure. I will attempt to do this by regressing the number of loans that have gone into default for any cycle on a

*Suny Oswego Applied Mathematical Economics Student, Senior

host of measures designed to assess each individual's ability to pay back the loans. I am interested to see if my logical predictions of what characteristics will cause an individual to fall into default will be proven accurate. The results of this analysis could potentially help microfinance institutions make more-informed decisions regarding who to provide loans to, in order to minimize their risk factors while still generating profit. Moreover, I am intrigued to prove or disprove the theory that females make better loan candidates than males. On one hand, it does seem logical that their business revenues will be reinvested within the family, spawning more consumption throughout the community and thus boosting incomes even more. But will the increase in disposable income stemming from successful businesses affect both genders of borrowers the same? Will this affect their comparative ability to repay loans?

DATA BACKGROUND

The data used for this analysis was pulled from a cross-sectional survey of 2,800 (randomly-chosen) impoverished households in Hyderabad, Andhra Pradesh, India. The survey was initiated by the Spandana microfinance company and was conducted in 2005. It spanned across 120 slums distributed throughout the area that were viewed as new market potential. The data was collected provisional on having an adult woman as a permanent resident in the household, making this sample a fitting choice for this regression. Although information was collected on all members of the household, I will use only the 1,891 observations aligning with the heads of each household. My rationale for this is based upon the fact that an analysis of ability to repay loans is contingent on an individual's credentials to receive loans. Thus, I have eliminated the children and those individuals in each household whose names do not appear on the loans. The original collection and analysis of this data was done by Abhijit Banerjee, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan in their paper entitled: "Measuring the impact of microfinance in Hyderabad, India." (Banerjee, Duflo, Glennerster, and Kinnan, 2008)

ALTERATIONS ON THE DATA SET

Some formatting changes were made on the original data set in order to be used in this regression analysis. As mentioned, only observations concerning the characteristics of the household heads were desired. Some information concerning the remaining members of the households was still deemed necessary however, so I have reformatted the individual observations of those members so they can be accounted for as well. For instance, the remaining number of observations beyond each household head was calculated in order to include a measure of the total number of residents per household in the regression. Children were accounted for in a measure of whether or not any of the remaining individuals were of or below the age of sixteen. Additionally, some of the household heads were recorded as maintaining two jobs, and total earnings in the month preceding the survey were given for each. These were combined for simplicity purposes, adding zero to the first measure for each individual that did not work a second job. Individual loan amounts were totaled, although the actual number of loans both past and present was extracted into its own category. Finally, for every individual, the number of loans that they had let slip into default for any one or more payment cycle(s) was quantified in addition to the existing dummy variable which simply represents whether or not that individual had ever made late payments.

INDEPENDENT VARIABLES AND DESCRIPTIONS

TABLE 1: Independent Variables

Variable Name	Variable Description
Age	Age of head of household (HH) at time of survey
BankSavingsAcco	Does HH have bank or savings account? (1=yes)
Childrenle16	Are there children 16 or under living in the household? (1=yes)
College	Is the HH's highest level of education college? (1=yes)
DeathsinthePast	Were there any major deaths in the family within the past year? (1=yes)
Employment	Is the HH employed at time of survey? (1=yes)
Gender	Gender of HH (1=male)
HHConditionPoor	Does the HH consider the household to be poor? (1=yes)
HS	Is the HH's highest level of education high school? (1=yes)
InsurancePolicy	Does the HH have some type of insurance policy? (1=yes)
LateRepayments	Have there been any late repayments for any cycle of the loan history? (1=yes)
Literacy	Is the HH literate? (1=yes)
LossofJobBusine	Has there been any loss of employment or closed businesses? (1=yes)
MaritalStatus	Is the HH married? (1=yes)
NuLateRepayment	Number of loans that went into default for any cycle
NumberinHH	Number of individuals living in the household
NumberofLoans	Number of loans (past and present)
PostGrad	Is the HH's highest level of education post-graduate studies? (1=yes)
RupeesonSickness	Any significant (500+ Rupees) expenditures on sickness? (1=yes)
TotalEarnings	Total earnings from job(s) in the past month (in Rupees)
TotalinLoans	Summation of all loans, both present and past (in Rupees)
**A dummy variable for "No education" was excluded to avoid perfect multicollinearity with the constant term*	

I have chosen to begin with these selected variables for this analysis because they will be good indicators of each individual's ability to repay loans. For instance, as the variable 'Childrenle16' increases I would also expect to see an increase the amount of loans allowed to go into default after factoring in the extra expenses, both time away from work and dollars spent, associated with caring for young children. Likewise, a household that had experienced any recent deaths or expensive sicknesses in the family would likely be facing steep and unexpected costs that might inhibit their ability to maintain a constant repayment schedule. The 'Gender' variable is especially important to this regression as it will help to explain my second hypothesis which was extracted from common thought: in the eyes of the microfinance institution, women are better borrowers than men. 'HHConditionPoor' is an interesting addition to this regression as it gives a reflective measure of the household's view of its own condition. This variable could potentially a proxy for how consumer (borrower) confidence affects repayment structure. Other variables like education which is represented by a series of dummy variables, insurance policies, literacy, marital status, and earnings all are expected to decrease the amount of loans allowed to default.

A Poisson regression will be used to model the finite number of possible outcomes for this regression. It is commonplace that a dependent variable with possible values that take on a maximum mean of [around] 10 could be subject to a Poisson distribution. In this case, it is possible to accurately model the data with a log model where it is assumed that the mean and variance of the distribution are equal. However, this occurrence does not often occur in practical situations. Instead, the variance is frequently greater than the mean of the data. This predicament violates the aforementioned assumption of the Poisson regression and is called “overdispersion.” To correct this, the data can be fitted to a negative binomial regression which does not assume that the mean is equal to the variance. By observing the spreadsheet of organized data, I expect that this will be the case in this situation. The characteristics of the observations, particularly those that have experienced some amount of delayed repayments, vary greatly throughout the observations.

RESULTS

1. Model I: Poisson Regression

Poisson, using observations 1-1891
Dependent variable: NuLateRepayment

	Coefficient	*Std. Error*	*z*	*p-value*	
const	0.0408765	0.179821	0.2273	0.82018	
Gender	0.045349	0.154405	0.2937	0.76899	
Age	-0.00961331	0.00248117	-3.8745	0.00011	***
Literacy	0.0798282	0.262563	0.3040	0.76110	
HS	-0.29521	0.261664	-1.1282	0.25923	
College	-0.71704	0.329672	-2.1750	0.02963	**
PostGrad	0.193594	0.440441	0.4395	0.66027	
MaritalStatus	-0.0642675	0.152882	-0.4204	0.67421	
NumberinHH	0.00841962	0.0143421	0.5871	0.55717	
Childrenle16	-0.0632752	0.0641165	-0.9869	0.32370	
Employment	-0.186901	0.0810885	-2.3049	0.02117	**
TotalEarnings	-1.5538e-05	1.06966e-05	-1.4526	0.14633	
DeathsinthePast	0.276254	0.073006	3.7840	0.00015	***
RupeesonSicknes	-0.00815229	0.0473261	-0.1723	0.86324	
LossofJobBusine	0.0857083	0.0763755	1.1222	0.26178	
NumberofLoans	0.281673	0.0103613	27.1851	<0.00001	***
BankSavingsAcco	-0.0447815	0.0549951	-0.8143	0.41548	
InsurancePolicy	-0.0338688	0.0587639	-0.5764	0.56438	
HHConditionPoor	-0.158515	0.0592095	-2.6772	0.00742	***
TotalinLoans	8.87748e-07	2.48202e-07	3.5767	0.00035	***

Mean dependent var	0.986779	S.D. dependent var	1.496145
Sum squared resid	3301.215	S.E. of regression	1.328312
McFadden R-squared	0.137362	Adjusted R-squared	0.130410
Log-likelihood	-2482.004	Akaike criterion	5004.008
Schwarz criterion	5114.905	Hannan-Quinn	5044.842

Overdispersion test: Chi-square(1) = 49.347 [0.0000]

At first glance, many of the predicted explanatory variables discussed above appear to be significant. College education, employment, and low confidence (considering the household to be of poor status) all appear to significantly lower the occurrence of default. According to these results, both the number of current outstanding loans and a recent death in the family will have positive effects on the amount of loans allowed to enter into a defaulted state. However, the combination of these independent variables does not seem to explain much about the distribution of this data set (13.7%). The overdispersion test for this regression also fails because the chi-square test-statistic is considerably higher than the critical value at this degree of freedom. This suggests that the data could in fact be better modeled by a different distribution. As discussed, a common solution to this issue is to assume that the data follows a negative binomial distribution where the conditional mean does not necessarily equal the variance.

2. Model II: Negative Binomial Regression

Negative Binomial, using observations 1-1891
Dependent variable: NuLateRepayment
Standard errors based on Hessian

	Coefficient	*Std. Error*	*z*	*p-value*	
const	-0.342519	0.236018	-1.4512	0.14671	
Gender	0.0729121	0.187621	0.3886	0.69756	
Age	-0.00552375	0.00314481	-1.7565	0.07901	*
Literacy	0.0728526	0.337105	0.2161	0.82890	
HS	-0.255206	0.335866	-0.7598	0.44735	
College	-0.643164	0.404684	-1.5893	0.11199	
PostGrad	0.306813	0.578032	0.5308	0.59556	
MaritalStatus	-0.0381855	0.185558	-0.2058	0.83696	
NumberinHH	0.000178903	0.0185031	0.0097	0.99229	
Childrenle16	-0.0734548	0.0802718	-0.9151	0.36015	
Employment	-0.116608	0.104541	-1.1154	0.26467	
TotalEarnings	-1.75844e-05	1.30769e-05	-1.3447	0.17872	
DeathsinthePast	0.303158	0.0989369	3.0642	0.00218	***
RupeesonSicknes	0.00629255	0.0602375	0.1045	0.91680	
LossofJobBusine	0.177162	0.102753	1.7242	0.08468	*
NumberofLoans	0.335251	0.0189512	17.6902	<0.00001	***
BankSavingsAcco	-0.0248117	0.0694219	-0.3574	0.72079	
InsurancePolicy	-0.0442324	0.0747345	-0.5919	0.55394	
HHConditionPoor	-0.171867	0.0750022	-2.2915	0.02193	**
TotalinLoans	6.45799e-07	5.82453e-07	1.1088	0.26754	
alpha	0.484328	0.0599961	8.0727	<0.00001	***

Mean dependent var	0.986779	S.D. dependent var	1.496145
Log-likelihood	-2387.722	Akaike criterion	4817.444
Schwarz criterion	4933.886	Hannan-Quinn	4860.320

The Pearson Goodness-of-Fit test in this case states that if the distribution of the dependent variable is significantly different than the Poisson distribution, then the reported “alpha” in this output should also be statistically different from zero. By comparing the p-value for “alpha” with a threshold value for type I error of 0.05, it can be concluded that the positive coefficient of "alpha" is in fact significant and different from zero. Thus, the negative binomial regression is appropriate and accurate in fitting this data set. Since this is the case, we can rely on our z-scores to be valid, consistent, unbiased, and BLUE predictors. The estimates for age and college-level education no longer appear to be significant in predicting late payments but the recent loss of employment or closure of business for a family seems to be more explanatory in this model than in the last. Further testing for misspecification of this model was performed and their results are included below. An examination of the variance inflation factors suggests that there may be multicollinearity between high-school-level education and literacy. This is both logical and noteworthy as it may be a testament to the efficiency of the school systems even in the most impoverished villages. It suggests that a high school education will almost always produce adequate literacy.

3. **Model II: Negative Binomial Regression, VIF**

Minimum possible value = 1.0	
Values > 10.0 may indicate a collinearity problem	
Gender	3.504
Age	1.313
Literacy	36.265
HS	36.708
College	5.027
PostGrad	1.472
Marital Status	3.500
NumberinHH	1.140
Childrenle16	1.141
Employment	1.355
TotalEarnings	1.266
DeathsinthePast	1.034
RupeesonSicknes	1.043
LossofJobBusine	1.037
NumbrtofLoans	1.125
BankSavingsAco	1.178
Insurance Policy	1.107
HHConditionPoor	1.124
TotalinLoans	1.205

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple correlation coefficient between variable j and the other independent variables

4. **Model II: Negative Binomial Regression, Likelihood Ratio Test**

Test for omission of variables -

Null hypothesis: parameters are zero for the variables HS

Test statistic: (-2)*[(-2388.004)-(-2387.722)] = 0.564

where $\ln(L_R)$ = log likelihood from Model III *(Next page)*

X^2 Chi-Square Critical Value: 3.8414

After executing a likelihood ratio test, it can be determined that the estimated parameter for high school education is not significantly different from zero so it can safely be dropped from the model without inhibiting the efficiency of the remaining estimators. The justification for aiming to eliminate this variable is to reduce the slight but present negative effects of multicollinearity. In doing so, I have estimated a final version of the model that best predicts the effects of the listed explanatory variables on the number of loans incurring late payments. The final model fits all assumptions of the negative binomial regression and may accurately predict the theoretical answers to my hypotheses. The estimated coefficients (reported below) of the independent variables are the expected change in the log of the expected count of loans in default, given a one-unit change in each variable. They can be exponentiated in order to clarify their interpretation. This procedure allows these changes to be expressed as percentages affecting the expected count of loans with late payments.

5. Model III: Negative Binomial Regression, correcting for collinearity

Negative Binomial, using observations 1-1891
Dependent variable: NuLateRepayment
Standard errors based on Hessian

	Coefficient	*Std. Error*	*z*	*p-value*	
const	-0.339908	0.236089	-1.4397	0.14994	
Gender	0.0740328	0.18764	0.3945	0.69318	
Age	-0.0055653	0.0031456	-1.7692	0.07685	*
Literacy	-0.178708	0.0643932	-2.7753	0.00552	***
College	-0.390441	0.230878	-1.6911	0.09082	*
PostGrad	0.558965	0.47377	1.1798	0.23807	
MaritalStatus	-0.0374916	0.185579	-0.2020	0.83990	
NumberinHH	0.000490864	0.0185009	0.0265	0.97883	
Childrenle16	-0.0732942	0.0802951	-0.9128	0.36134	
Employment	-0.120099	0.104463	-1.1497	0.25028	
TotalEarnings	-1.77398e-05	1.3095e-05	-1.3547	0.17551	
DeathsinthePast	0.301192	0.0989443	3.0441	0.00233	***
RupeesonSicknes	0.00649151	0.06026	0.1077	0.91421	
LossofJobBusine	0.174638	0.102748	1.6997	0.08919	*
NumberofLoans	0.335255	0.0189653	17.6773	<0.00001	***
BankSavingsAcco	-0.0272324	0.0693727	-0.3926	0.69465	
InsurancePolicy	-0.042222	0.074713	-0.5651	0.57199	
HHConditionPoor	-0.172826	0.0750255	-2.3036	0.02125	**
TotalinLoans	6.55567e-07	5.81787e-07	1.1268	0.25982	
alpha	0.4855	0.060014	8.0898	<0.00001	***

Mean dependent var	0.986779	S.D. dependent var	1.496145
Log-likelihood	-2388.004	Akaike criterion	4816.007
Schwarz criterion	4926.905	Hannan-Quinn	4856.842

With this knowledge, it can be concluded that one-unit increases in age or literacy will decrease the expected count of loans in default by 0.55% and 16.36%, respectively. Moreover, a college-level education or low consumer confidence (reporting one's own household as poor in financial condition) could decrease the expected count by 32.32% and 15.87%, respectively. Conversely, a recent death in the family or a loss of employment (or closure of a business) can be expected to increase the number of loans an individual will let fall into default at some point during repayment to increase by as much as 35.15% and 19.08%, respectively. Finally, with each additional outstanding loan, we can expect to observe a 39.83% increase in the total number of defaulting loans.

The implications of these results are substantial; evidence has been provided for both of my hypotheses. First, the effect of gender on the ability to repay loans is statistically insignificant here meaning that the expected count of loans in default does not vary between men and women. This suggests that the status quo of targeting women with microfinance under the assumption that they will be less risky may not reward creditors with the same magnitude of gains as commonly expected and may not be based on factual evidence. Larger data sets could be examined to confirm these results. Second, a SAS analysis of earnings based on gender proves that there is a significant difference between the two levels but that there is no difference between household financial conditions (based on the variable 'HHConditionPoor '). This implies that although incomes differ between genders (in favor of men, not women), the impact on the household does not; negating the theory that microloans benefit women more than men.

The two outputs shown below are both analyses of levels conducted with SAS. Their outcomes imply that not only should individuals be assessed for their ability to repay loans but so should the villages that they reside in. The first output is an analysis of how the total amount of loans in each individual's name, expressed in Rupees, differs between slums. The results show that there is a significant difference in total amount of loans among the villages. This is evidence that some villages may be made up of good communicators, sharing their successes with the loan programs through word of mouth. Microfinance institutions could use this information to take advantage of cheap (or near-free) advertising and marketing opportunities. The second output conveys that the number of loans that have fallen into default differs significantly by village as well. Thus, the same institutions could exploit the data presented here to better analyze which villages would be less risky (and more profitable, ceterus paribus). This evidence taken with the results presented earlier in this paper may prove this tactic to be more beneficial to the overall status of the institution than targeting women.

The hypotheses presented in this paper have thus far been proven to be correct; based on the results, loan defaults can be accurately modeled through regression analysis, women do not necessarily make less-risky borrowers than their male counterparts, and household conditions do not drastically improve under the sole stipulation that the signer on the loan is female. The inclusion of additional explanatory variables, larger samples of households under analysis, or the exploration of other possible models could further prove or disprove these results.

6. SAS Analysis of Levels: Slum

The SAS System

The ANOVA Procedure

Dependent Variable: condition

Number of Observations Used: 1891

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.2608350	0.2608350	1.47	0.2248
Error	1889	334.1326076	0.1768833		

The SAS System

The ANOVA Procedure

Dependent Variable: earnings

Number of Observations Used: 1891

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	441922072	441922072	40.02	<.0001
Error	1889	20858209061	11041932		

7. SAS Analysis of Levels: Gender

The SAS System

The ANOVA Procedure

Dependent Variable: Total

Number of Observations Used: 1891

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	119	968529199593	8138900836.9	1.45	0.0016
Error	1771	9.9666682E12	5627706502.1		

The SAS System

The ANOVA Procedure

Dependent Variable: NoLate

Number of Observations Used: 1891

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	119	855.168501	7.186290	3.77	<.0001
Error	1771	3375.500986	1.905986		

REFERENCES

Abhijit Banerjee; Esther Duflo; Rachel Glennerster; Cynthia Kinnan, "Measuring the impact of microfinance in Hyderabad, India", http://hdl.handle.net/1902.1/11389 UNF:5:7llipBUQ4zNQHjfYYJVqwA== MacArthur Data Consolidation Project [Distributor] V5 [Version] http://ifmr.ac.in/cmf/?page_id=440