How demographics contribute to mortgage default: using binary logistic regression to draw inferences

↳ By DLPM Rathnasingha and Ali Khatibi

1. Introduction

Housing is an indispensable human need. It is predictable as a basic need for shelter. It is apparent that habitable housing contributes to the health, efficiency, positive social behavior and general welfare of the populace, which uplift the living conditions of the community. In present society, housing has a value beyond the essential need for human shelter. As an upshot many agencies have moved to work in this field. This has made for a rapid transformation in the household sector in Sri Lanka. Recent studies expressed concern that substantial numbers of bank customers, who enter into complicated financial contracts such as mortgages, are financially illiterate and fail to understand crucial factors in decision-making, Engin (2011). Further, for most governments the availability of sufficient but basic housing for all is often stated as a priority for enhancing the social needs of the society. For a typical house-owner, the house is a major asset in his portfolio and for many households the purchase of a house represents the largest (and often only) lifelong investment and a store of wealth, Goodman (1989). Housing represents a large proportion of a household's expenditure and takes up a substantial proportion of lifetime incomes. It has been noted by Rose (2011) that for most Americans, the purchase of a home is the greatest and most important financial decision made during their life time.

A central purpose of financial markets is to redistribute risk in an optimal manner. However, external influences such as local currency depreciation, oil and gold price increases plus some internal factors, including higher levels of interest rates, have made the financing of housing a major challenge for the country. It needs attention at a policy level. The provision of housing services depends mostly upon a well-functioning and affordable housing finance system. The housing finance process is very complicated and the beneficiary has to meet a set of difficult criteria which make it harder to meet the requirements for a housing loan. Further, the interest rates on loans are unaffordable for the general working class of people and they went up to nearly 20% per annum in the recent past. This rate has shown vast fluctuations from 20% to 12% in the period 2009-2012 indicating the instability of the market. This rate fluctuation creates a number of issues for lenders and for borrowers in relation to their financial management. It has been mentioned by Vandell et al. (1985) that understanding the forces that drive mortgage default is a necessary prerequisite to appropriate pricing of default in the mortgage market. This study is focused on the demography of defaulting borrowers and provides valuable insight into the demography of mortgage borrowers.

2. Current mortgage loan environment

Sri Lanka has a population over 20.2 million and the density of population is 326 persons per square kilometer. However, annual housing demand for new houses is estimated at around 80,000 to 100,000 units in the country. The size of the housing finance sector in the country is 6% of Gross Domestic Product (GDP) though there are no records on a substantial number of owner-occupied and self constructed houses. Empirical research on housing finance in Sri Lanka is scarce due to the paucity and cost of relevant data. The mortgage default problem has been looked at by a number of researchers from different angles but mainly in Europe, although the United States has experienced a tremendous wave of residential mortgage foreclosure in recent years W. Scott (2010). One must look for indicators of risk for the housing finance sector in Sri Lanka. It was reported in some of the industry reports such as the Lanka property web on 14/06/2012 that the Sri Lanka overall house sale price is 17.11 million following a 4.61% down turn in price. A recent precipitous fall in the market value of the housing stock has created an ongoing sub-prime crisis in the sector. It was pointed out by Kim (1997), Manda et al. (2011) that in the absence of a well-functioning housing finance system, market-based provision of housing would therefore be lacking. This focuses the attention of researchers, policy makers and government decisions makers on addressing this situation to create affordable housing finance solutions for the country that go beyond lucrative business opportunities. Despite several attempts by the government of Sri Lanka to address this state of affairs, there are no approaches that have achieved sufficient scale and hundreds of thousands remain in need. How successful the housing finance system in Sri Lanka becomes will depend to a significant extent on how well it is integrated with, and forms part of, the overall financial sector. Under present conditions this is questionable as there is no secondary market for mortgage-backed securities in the country.

Demographic changes such as the growing number of working women, the expansion of nuclear families and a growing urban population are driving housing demand in the country. In societies like Sri Lanka, where social housing is on the priority list of government, housing affordability would have to be looked at from several angles, including the point of view of an individual's ability to raise the money needed to meet the cost of their housing needs. In the recent past several institutions have made efforts to improve access to loan financing for housing in the country. However, it is believed that there is an enormous unmet demand for housing finance from the low-income segment of the society.

The first source of funding for housing for many individuals is the proceeds of their income. This is often the cheapest source because there is no payment of extra cost in the form of interest and no other requirements to fulfill such as documents, guarantees etc. The second funding source is financial arrangements such as mortgage finance. It can be observed that there are government, private and non-government organizations providing finance facilities for housing in the country. Banks in Sri Lanka have being playing a vital role in providing credit to the housing sector for a few decades. However, Piyasiri (2006) states that the mortgage payment collection ratio for housing finance institutions in Sri Lanka is estimated to be in the range 80 to 95%. Further he identified that in certain cases the proportion of non-performing loans is around 20%. This is a major problem for finance suppliers to the sector although they have security and guarantees.

The two main players in the housing finance sector in Sri Lanka are the National Savings Bank (NSB), fully owned by the Government and the Housing Development Finance Corporation (HDFC) which is a listed company although the main share holder (51%) is the Government. During the year 2010, NSB granted housing and property loan balances of 33,011,819 (Rs' 000) which represents an increase over the 2009 balance of 28,339,675. This is equivalent to a 16.48% increase in housing and property lending by the institution. Comparatively the loan loss provision balance on total loans at the beginning of the year 2010 was 542,807 which had increased to 646,176 at the year end. However, at the beginning of year 2009 loan loss provision was only 365,907 Rs making a 19.94 % of increase year-on-year. Importantly the balance of loan loss provision at the beginning of 2008 was only 180,161. (NSB Annual Report 2010, 2009)

As per the District-wide summary of HDFC at April 2012, there were 70,380 loans granted amounting to Rs. 14,215,015,566.58 and 36,713 of these loans were under EPF[1] security which is 52.16% of the total housing loan portfolio, with a balance of 33,667 loans under other security types. There were a total of 24,178 loans making 34.35% of total loans which had been identified as in default and out of them 19, 588 loans in default were granted under EPF security amounting to 81.01% of total loans in default. There were only 4,590 default loans out of non EPF security loans granted by HDFC. At the branch level, for the total number of loans disbursed the minimum default rate was 0.16% while the maximum was 57.57% which indicates a huge variation between branches. However, excluding EPF security loans this range is only between 1.22% at the minimum and 19.91% at the maximum. Out of the total housing loans contracted 52% were allocated to the western province and 11% to the central province while other provinces reported less than 10% each.

3. Model of risk assessment

In this study, six independent variables are used: age, gender, marital status, number of dependants, employment status and credit period. Age is specified by the number of years, while gender is specified as male or female, marital status as married, unmarried or other, number of dependents as the number, employment status as government, private or other, and credit period as the total number of monthly installments to pay for the period of the loan. The dependant variable, default, was assigned dummy values equal to 0 if the loan defaulted and 1 otherwise. A number of techniques have been used to estimate the likelihood of default over time, Dennis et.al (2005). The final Binary Logistic regression model [BLR] is expected to yield a prediction of the probability of default on loan repayments, Wilson (2003). In this study BLR has been employed to gauge the probability of default and this can be used when the dependent variable is a dichotomy/division of some kind into two groups.

4. Survey of mortgage loan customers

Understanding of the determinants of default has been severely constrained by the non-availability of suitable data, Campbell at el. (1983). Today there are no publicly available databases in relation to individual outstanding mortgages and default in Sri Lanka. As a result this researcher requested access to data compiled by one major player in the market. This institution[2], under government ownership seeks the following requirements if prospective borrowers are to be eligible for a mortgage house finance facility. They include both qualitative and quantitative factors to evaluate the credit worthiness of their customers when evaluating the loan application. The applicant has to be a permanent employee or a tax payer and only a maximum 64% of monthly salary/income can be used for monthly loan settlement. The original data set of the institution included information on the above-mentioned variables for 4,237 borrowers. However, it was observed there were some missing variables in many cases and it was able to provide complete data on 1,000 loan holders, out of which 500 from the total database are borrowers in default, for analysis. All the above loans were for housing and information on mortgages for other purposes, such as commercial loans, has not been entered into the database.

5. Descriptive statistics on the data base

The mean age of the sample is 45.33 years (appendix 01) with a minimum age of 20 years and a maximum of 70 years. There are 649 males making 64.9% of the total sample and 351 females. Married individuals in the sample are 957 which comprise 95.7% of the total and unmarried 43. With regard to employment status, 950 individuals which are 95% of the sample are government workers with only 36 private sector individuals making 3.6% of the sample with 14 individuals (1.4%) under other employment categories. The number of dependents was a minimum of 0 and the maximum is 5 with a mean of 1.40 which is treated as one dependant. However, 30% of the sample does not have dependents while 19.3% have only one dependant and 35.1% had 2 dependents at the time when the mortgage loan application was approved. Among 295 of loan holders with 0 dependants the mean credit period is 165.11 months (nearly 15 years) with a minimum of 18 months and a maximum of 360 months. The average loan term is 164.15 months which is equivalent to nearly 15 years, and there are only 160 loans out of total sample with a term of less than five years.

According to gender and default cross tabulation (Table 01:A.), 53.6% of males defaulted while only 43.3% of females defaulted. Based

[1] EPF is the abbreviation for Employees Provident Fund. Its primary aim is to provide a measure of security for old age retirement to its members. Employer and employee contribute to the fund as per the directions of the labor commissioner.

[2] The researcher is not at liberty to mention the name of the institution for reasons of confidentiality.

[3] The odds ratio is a relative measure of risk, telling us how much more likely it is that someone who is exposed to the factor under study will develop the outcome as compared to someone who is not exposed.

on the odd ratio[3], the males are 1.5 times more likely to default compared to females. Among those government employees who defaulted 55% were males and 43.3% were females but in the private sector male and female borrowers do not show a difference in the rate of default. Conversely, in the "other" sector no males have defaulted on their mortgage loan while 66.7% females have defaulted and 17.9% of unmarried males and 40% unmarried males have defaulted.

However, comparison between default and marital status to see how likely married borrowers are to default with use of an odds ratio[4] of 2.938 indicates that married loan holders are almost three times more likely to default compared to unmarried loan holders While among married borrowers 51.1% (Table 01: B.) have defaulted, and only 25.6% of unmarried borrowers have defaulted on their loans. Further, among defaulted borrowers 55.2% (Table 01: E.) were married males while 43.5% of married female have defaulted their loans. The Pearson's chi-square[5] value of 10.71 (appendix 06) on default and marital status with a p-vale 0.001 and minimum expected count of 21.50 indicates that there is an association between marital status and default.

Cross tabulation on the number of dependants and default (Table 01: C.) shows a gradual increase in default amongst those with higher numbers of dependants. Among the total sample 46% of defaulting borrowers had 0 dependants while default levels were 46.1%, 51.4%, 55.7%, 76.5% and 100% for borrowers with 1,2,3,4 and 5 dependants respectively. There were only 4 loan holders with 5 dependants and they all defaulted on their loans but one cannot apply further analysis as the numbers are too small to be statistically valid. The Pearson's chi-square statistic on default and number of dependants is 13.39 (appendix 07) with a p-value of 0.02, hence there is an association between default and the number of dependants.

The data on the employment status of loan holders shows (Table 01:D.) that among government sector loan holders 50.9% has defaulted on their loan while only 38.9% of private sector loan holders have experienced default of their loan and among other sector loan holders only 14.3% have defaulted on the loan. Chi-square statistics on default and employment status 9.26 (appendix 08) under a p-value of 0.01, show there is an association between employment status and default.

6. Explanation of the risk assessment model

The output of the model (BLR) $Z = \alpha + \beta_1 I_1 + \beta_2 I_2 + \ldots\ldots\ldots + \mu I$ is expected to yield the prediction of the probability of default on housing loan repayments. Values for the independent variables (factors of default) can be substituted in the formula to calculate an estimation of the likelihood of default. Then, using the estimated equation calculating the observed value for Z, the value of the Logit model is transformed using a link function to obtain the probability of the event (default) occurring[6]. The BLR model can work by applying β values (Table 02) of independent variables for default risk in this study and it is not necessary to discuss deeply the statistical/ mathematical procedure of BLR involved.

Equation Z = -5.011 + 0.005(Age) + 0.566(Gender) + 1.203(Marital status) - 0.129(Number of Dependants) + 0.729(Employment Status) + 0.013(Credit Period) + μ.

The results of the BLR model show that marital status is most significant (β = 1.203 and p = 0.00) among the variables and another influencing factor for defaulting is employment status (β = 0.7293 and p = 0.00) while gender (β = 0.566 and p = 0.00) also has a considerable weight. One other contributing factor is the number of dependants (β = -0.129 and p = 0.03). Age and credit term is indicated to be least important for default on loans and age is not significant as the p-value is more than 0.05.

With insertion of values on independent variables one can apply the model and the following demonstrate application of them to two cases in the sample.

Z = -5.011+0.005 * (45) + 0.566 * (1) + 1.203 * (1) - 0.129 * (0) + 0.729 * (1) + 0.013 * (120)
Z = -0.728

Table 1: E. Gender * Default * Marital status cross tabulation

Marital status				Default	
				Yes	No
Married	Gender	Male	Count	343(55.2%)	278(44.8%)
		Female	Count	292(43.5%)	380(56.5%)
Unmarried	Gender	Male	Count	5(17.9%)	23(82.1%)
		Female	Count	12(40.0%)	18(60.0%)

Table 2: Results of BLR

		β	S.E.	Wald	Sig.[7]
Step 1[a]	Age	.005	.005	1.085	.297
	Gender	.566	.145	15.313	.000
	Marital status	1.203	.376	10.218	.001
	No of dependants	-.129	.062	4.319	.038
	Employment status	.729	.266	7.489	.006
	Credit period	.013	.001	99.653	.000
	Constant	-5.010	.647	59.959	.000

a. Variable(s) entered on step 1: Age, Gender, Marital status, No of dependants, Employment status, and Credit period.

[4] "The odds" is the ratio of the probability that the event of interest occurs to the probability that it does not. This is often estimated by the ratio of the number of times that the event of interest occurs to the number of times that it does not.

[5] The chi-square test of association is used to discover if there is a relationship between two categorical variables. A large chi-square, indicated by a small p-value, is taken as an evidence of association.

[6] The link function is as follows: $P[event] = \frac{e^z}{1+e^z}$

[7] S.E. – Standard Error, Wald –Wald statistics , Sig. - Significance

Next the above Z value is applied to the link function and one obtains the probability of the event (default) occurring as shown below.

$$P[event] = \frac{e^{-0.728}}{1+e^{-0.728}} = 0.3256$$

By applying the link function and assuming a 50:50 of cutoff is used, the probability of this respondent's loan defaulting is 33% (0.3256 ≈ 33%).

The following is the second case which applies values for independent variables from the sample.

Z = -5.011+ 0.005*(59) + 0.566*(1) + 1.203*(1) - 0.129*(0) + 0.729*(1) + 0.013*(249)
Z = 0.902

And then apply link function.

$$P[event] = \frac{e^{0.902}}{1+e^{0.902}} = 0.7113$$

Using the BLR model the probability of this respondent's loan default is 71% (0.7113 ≈ 71%). Then in the first scenario it has a less chance of getting default the loan. However, in the second scenario the possibility of default is high. The accuracy level of the model is crucial and so it was applied to the model for the whole base of the sample and produced results for prediction accuracy as shown in table 03 below.

The accuracy of prediction (Table 03) is at an acceptable level which is better than a random prediction with an overall accuracy level of 67.5% and therefore this model provides improved explanations on predicting the probability of a housing loan being in default using six independent variables.

7. Observations

The current mortgage market is becoming full of challenges with new practices and rapid changes of customer demographics which make for continuous transformation of the mortgage industry. However, within these changes, the sample in this study shows that the number of dependants is not a decisive factor in default as the majority of the sample does not have dependants at the moment of obtaining the mortgage facility though this possibly will change over the term of the loan as borrowers are blest with children. In addition, householders may borrow too much at a high rate without realizing the future risks. This indicates the need for updated data on the age of the mortgage applicant (the age at which they default) as a determinant of mortgage loan default. Virtually four fifths of the sample is made up by males, indicating that males are more likely to engage in housing finance facility arrangements in the market.

The large majority of the sample is made up of state of government workers, proving that a large proportion of mortgage loans granted are obtained by them. On the other side this highlights the difficulty of obtaining housing finance facilities for people working in other sectors or in meeting the criteria to gain access to a mortgage loan facility in the current market. The difficulty of meeting mortgage criteria hinders access to credit by low-income citizens. It implies that it is hard for other sector individuals to obtain a mortgage facility. More importantly, there are no government programs relevant to urban residents working in the informal sector who have low incomes and who are unable to apply for loans from the formal sector. This is suggests another market segment for microfinance market players to provide housing finance facilities though the current scale of their activities is very small.

The BLR model developed and tested here is a useful tool for lenders in the industry to gauge the risks of default and implement necessary strategies to minimize the risks. The accuracy level of the predictions with the model as developed is acceptable. However, it is realized that there are a few more factors such as economic and location factors to be incorporated into the model and tested in relation to the level of default to make estimating the level of default risk more reliable.

Table 3: Default observed * Default predicted cross tabulation

Count				
		Default Predicted		Percentage correct
		YES	NO	
Default observed	Yes	363	137	72.6
	No	188	312	62.4
Overall Percentage				67.5

References

Campbell T.S., Dietrich J.K. (1983), The Determinants of Default on Insured Conventional Residential Mortgage Loans, The Journal of Finance, Vol XXXVIII, No 05, December, pp. 1569-1581.

Dennis G. and Peter N., (2005), Measuring the Default of Small Business Loans: a Survival Analysis Approach, Journal of Money, Credit, and Banking, Vol. 37, No. 05, October, pp.923-947.

District-wide summary (April 2012), Housing Development Finance Corporation (HDFC).

Engin Demiral. (2011), Financial Behavior and Demographic factors on Mortgage Decisions, International Research Journal of Finance and Economics, 74, pp. 207-217.

Goodman, A.C. (1988), An Econometric Model of Housing Price, Permanent Income, Tenure Choice, and Housing demand, Journal of Urban Economics, 23, pp. 327-353.

Kim, K. H. (1997), Housing Finance and Urban Infrastructure Finance, Urban Studies, Vol.34 No 10, pp. 1597-1620.

Manda, M.A.Z., Nkhoma S. and Mitlin D. (2011) Pro-poor housing finance in Malawi, Human settlement working paper aeries. June.

NSB Annual Report 2010, 2009.

Piyasiri, S. H. (2006), Home Loans- Recent trends in Sri Lanka, Housing Finance International, June 2006, pp.29-33.

Rose C. C., (2011), Qualifying for a home mortgage in today's mortgage environment, Journal of financial service professionals, March.

Scott Frame, W. (2010), Estimating the Effect of Mortgage Foreclosures on Nearby Property Values: A Critical Review of the Literature, Federal Reserve Bank of Atlanta-Economic Review Vol. 95, November 3.

Vandell Kerry D., Thibodeau T. (1985) Estimation of mortgage defaults Using Disaggregate Loan History Data, AREUEA Journal, Vol. 13, No 03, pp. 292- 316.

Wilson J.H. (2003), An Analytical Approach To Detecting Insurance Fraud Using Logistic Regression, Journal of Finance and Accountancy, pp. 1-15.

Appendices

SPSS Analysis output tables

Descriptive overview of data base

	Minimum	Maximum	Mean
Age	20	70	45.33
No of dependants	0	5	1.40
Credit term	18	360	164.15
Default	0	1	.50

Gender

	Frequency	Percent
Male	649	64.9
Female	351	35.1
Total	**1000**	**100.0**

Marital status

	Frequency	Percent
Married	957	95.7
Unmarried	43	4.3
Total	**1000**	**100.0**

Employment status

	Frequency	Percent
Government	950	95.0
Private	36	3.6
Other	14	1.4
Total	**1000**	**100.0**

Number of dependants

	Frequency	Percent
0	295	29.5
1	193	19.3
2	351	35.1
3	140	14.0
4	17	1.7
5	4	.4
Total	**1000**	**100.0**

Pearson Chi-Square on default and marital status

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.717[a]	1	.001
Likelihood Ratio	11.169	1	.001
N of Valid Cases	1000		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 21.50.

Chi-Square on default and number of dependants

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13.394[a]	5	.020
Likelihood Ratio	15.197	5	.010
Linear-by-Linear Association	8.407	1	.004
N of Valid Cases	1000		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 2.00.

Chi-Square tests on default and employment status

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9.262[a]	2	.010
Likelihood Ratio	10.059	2	.007
Linear-by-Linear Association	8.910	1	.003
N of Valid Cases	1000		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.00.