Evaluating the impact of demographic characteristics on residential mortgage default risk: Evidence from Lebanon

Ali Mahmoud Antar

Economics Department, Beirut Arab University, Lebanon.

(Email: ama935@student.bau.edu.lb)

ABSTRACT

Purpose: This study aims to examine the relationship between borrower's demographic characteristics and default risk in mortgage loans to help financial institutions develop more effective lending policies.

Design/Methodology/Approach: Cross-sectional data were elicited from randomly selected 6743 individual accounts from Lebanese housing banks. This study applied the binary logistic and stepwise regression models to analyze the dataset using the Stata statistical software. Model diagnosis is performed using the Hosmer-Lemeshow goodness of fit test, likelihood ratio test, model accuracy classification table and statistically significant test-ROC curve.

Findings: The findings revealed that there is a significant relationship between residential loan default risk and borrower's marital status, nature of job occupation, job economic sector, job location and loan purpose. The performance of the binary logistic regression analysis demonstrates the overall percentage who is correctly classified is 91.61%.

Conclusion: The log odds of default risk for widowed borrowers are about 90 percent higher than those of divorced borrowers and that of self-employed borrowers is about 54 percent higher than that of employed borrowers. Borrowers working in the banking and real estate sectors have lower default rates than borrowers working in other economic sectors. In addition, loans granted for renovation purposes have the lowest default rates compared to loans provided for purchase, under-construction and construction purposes.

Practical Implications: The empirical results help financial institutions to have early warning signals in detecting financial distress and to differentiate between a high- and low-risk group of borrowers, helping in the development of tailored risk mitigation strategies and adjusting the lending criteria.

Keywords: *Credit risk, Housing finance, Loan default, Risk management, Risk mitigation, Binary logistic regression, STATA software.*

1. INTRODUCTION

Lebanon has been facing a severe economic crisis since October 2019 which led to critical economic depression, high inflation, a skyrocketing unemployment rate and severe currency devaluation. The main reasons behind the Lebanese collapse are due to political instability, widespread corruption, regulatory practices, government debt, and long patterns of non-productive economy. The Lebanese pound has lost 98% of its value against the US dollar leading to a sharp decline in consumer purchasing power and standard of living. In addition, the Lebanese banking sector has been significantly impacted by the economic crisis and this has led to a sharp increase in non-performing loans since most salaries are paid in local currency. The capital control practiced by the Lebanese banks enhanced restricted access to foreign currencies and this led to an increase in default rates which reached alarming levels.

The overall soundness of the economy, financial institutions and households were all negatively affected by the rise of loan default rates. The Lebanese government issued a one-year regulation prohibiting banks and financial institutions from pursuing any legal action against borrowers facing financial troubles due to their failure to make

Nurture*: Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN*: 1994-1633/* ***Print ISSN****: 1994-1625*
DOI*: 10.55951/nurture.v18i4.901|* ***URL****:* www.nurture.org.pk
Publisher: *Nurture Publishing Group*

timely loan payments. Therefore, it is crucial to identify the factors that contribute to default risk and to develop effective strategies to mitigate such risk to avoid the unfavorable consequences of housing loans defaults.

There is a growing recognition of the need to examine segmented borrowers' demographic and categorical parameters that influence default risk and help to predict the risk of default not only at the loan origination stage but also during the loan's term to maturity and during economic distress and financial instability while previous studies have expensively explored quantitative factors affecting default risk in residential mortgage loans such as borrower credit score, loan-to-value and debt ratio which evaluate the risk profile of borrowers at the loan initiation stage.

The Lebanese regulatory system represented by the Lebanese Central Bank and the Lebanese Bank Control Commission requests from all banks in Lebanon to apply to the following financial ratios when evaluating the borrower risk profile. The mandatory lending criteria set by regulatory stated that the maximum monthly housing loan payment relative to income must not exceed 35 percent of the applicant's income. If the applicant has other retail loans such as car, educational or personal loans besides the housing loans, then the maximum monthly loan payment relative to family income must not exceed 45 percent of the applicant's family income. In addition, the lending policies also indicated that the maximum loan-to-value ratio shall not exceed 75 percent which means that the borrower shall have at least 25 percent of the mortgage price as a down payment. Furthermore, Lebanese regulations set by the Lebanese Central Bank impose limits on the amount that can be lent to clients to ensure that borrowers do not overextend themselves financially and that banks do not take on excessive risk. However, none of the lending criteria set by the central bank include any demographic parameters such as the borrower's age, marital status, number of dependents, existence of additional collateral, job economic sector whether borrower income is paid in local or foreign currency etc. This study aims to fill this gap by empirically investigating the impact of various demographic characteristics on the risk of default in residential mortgage loans. We explore how demographics and categorical parameters have a significant influence on the risk of default using data from 6743 individual housing loan accounts from Housing Finance Institutions in Lebanon for the period between the years 2005 and 2020. The importance of this paper is reflected by the fact that not only the borrower's categorical variables will be examined but also the sub-categorical variables will be included. The variables used in the current study are gender including three sub-categorical variables which are male, female and co-borrower married couples, marital status which is composed of the single, married, widow and divorced variables, borrower's job type variables including employed, freelancer and self-employed categories, job industry variables including banking, service, commercial, industrial, real estate, public and private sectors, type of loan requested by the borrower including purchase, construction, under-construction, and renovation loans and borrower's job location, the existence of dependence and the existence of additional collateral. The empirical results of the study are expected to provide valuable insights for financial institutions in developing more effective lending policies. In addition, the study seeks to inform better risk management practices and contribute to the stability and resilience of the housing finance industry.

Examining the impact of demographical parameters on default risk can assist lenders and financial institutions in identifying risky loans and therefore they can better select borrowers with less exposure to credit risk. In addition, identifying factors that lead to default risk also provides borrowers early signals of the factors that might affect their income negatively so that they can take corrective actions to avoid loan default and its negative consequences. Furthermore, it will also help policymakers to develop policies and regulations that better lead to a sustainable housing finance system and be able to develop effective strategies to mitigate this risk.

The assessment of demographic parameters helps in evaluating the patterns of the borrower's behavior. For instance, during the economic crisis and financial distress, default risk exposure for borrowers who have employed jobs is different than those who are freelancers. Moreover, different economic sectors are exposed to different degrees of risk of default. Furthermore, when the Lebanese economy suffers from currency devaluation, borrowers who work outside Lebanon or those who work in the private sector and earn a salary in foreign currency have a lower risk of default than those who earn a salary in Lebanese local currency.

This paper is presented as follows: Section two presents the literature review. Sections three and four present the econometrics model and methodology followed by data analysis and model development in section five. Section six presents the model diagnosis procedure followed by results and discussion and section 7 presents the conclusion.

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

2. LITERATURE REVIEW

The literature reviews have highlighted the significant impact of borrower demographic characteristics in predicting the likelihood of default risk in residential mortgage loans. Many research studies have shown that variables such as age, gender, educational level, income, nature of job occupation and marital status can affect default risk.

A study conducted by Goel and Rastogi (2023) revealed that younger borrowers often have a higher risk of default than older borrowers. This can be attributed to the lack of credit history and financial stability among younger individuals. Conversely, older borrowers might have a more established credit history and more equity and assets, which reduce the probability of default rates. This study suggests that age is a significant predictor in credit scoring models with default risk decreasing as an increase in borrower age and accordingly there is an inverse relationship between the risk of default and the age of a borrower.

In addition, gender has also been found to influence the risk of default even though the impact can vary according to the geographical region and cultural context. Some studies revealed that women have a lower risk of default than men due to different spending and saving behaviors. However, these trends are associated with countries and cultures related to whether women are income-independent or not (Goel & Rastogi, 2023). However, another study conducted by Munnell, Tootell, Browne, and McEneaney (1996) found that female borrowers and single borrowers had higher default rates suggesting that household structure and gender dynamics influence financial stability.

Furthermore, a study conducted in 2023 by Noriega, Rivera, and Herrera (2023) shows that higher educational levels of borrowers are generally associated with lower default risks. Educated borrowers tend to have better job opportunities, higher incomes, and greater financial stability which contribute to a higher financial ability to pay their monthly obligations. Many empirical studies recommended that education is a critical demographic factor that should be included when developing a credit scoring model.

Moreover, the borrower's job nature and job industrial economic sector are also significant predictors of default risk. Studies by Fisher and Gervais (2011) indicated that self-employed individuals and those working in volatile industries are more prone to default due to income instability. This supports the findings of Borzekowski and Cohen-Cole (2008) who noted that employment stability is a critical factor in mortgage performance.

Furthermore, geographical location and housing market conditions also play a crucial role. Goodman, Seidman, and Zhu (2013) found that properties in economically distressed areas have higher default rates highlighting the impact of local economic conditions on borrowers' ability to repay loans. Additionally, the type of mortgage product and loan terms can influence default risk. Foote, Gerardi, and Willen (2008) showed that adjustable-rate mortgages (ARMs) had higher default rates compared to fixed-rate mortgages, particularly in periods of rising interest rates.

Nevertheless, many studies show that marital status is one of the important demographic parameters that significantly affect the risk of default because it can influence financial stability and borrowing behavior. Married individual co-borrowers may benefit from dual income and share financial obligations and this potentially reduces the risk of default. However, single or divorced borrowers might face more financial distress and this increases their likelihood of default (Lee & Sohn, 2021).

Furthermore, many studies highlight the role of the existence of additional collateral in reducing credit risk. Collateral provides a buffer against potential loss where borrowers who provide additional collateral along with the main one which is the housing unit subject of the loan will be more cautious to pay their monthly liabilities to avoid losing both collaterals in case of default. Therefore, collateralized loans have a lower probability of default compared to non-collateralized ones (Jimenez, Salas, & Saurina, 2006).

Furthermore, many studies were conducted to examine the relationship between the borrower's job industry and the risk of default since the job industry can significantly influence the borrower's income stability and job security. Some industries are more sensitive to economic fluctuation, distress, and other factors that can affect the borrower's ability to repay the loan. Another study revealed that the job industry is a critical factor that should be included in a scoring model to predict default risk and that industries with higher volatility and lower job security such as construction and retail tend to have higher default rates (Banasik & Crook, 2005).

Moreover, job location can have a significant impact on default risk since every job location is associated with its regional economic conditions, employment opportunities, and cost of living. Borrowers in economically strong

Nurture: *Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN*: 1994-1633/* ***Print ISSN****: 1994-1625*
DOI*: 10.55951/nurture.v18i4.901|* ***URL****:* www.nurture.org.pk
Publisher: *Nurture Publishing Group*

regions with low unemployment rates are less likely to default and regions with higher unemployment rates and declining property values revealed an increase in default rates emphasizing the need to consider job location in credit risk models (Foote et al., 2008).

In addition, many studies examine the impact of housing location on the default factors. The results revealed that property location has a significant influence in determining default risk and those locations associated with negative neighborhood characteristics, weak local economic conditions, proximity to amenities and high crime rates that increase homeowner distress lead to a high risk of default. For instance, a study was conducted to examine the influence on default risk of borrowers where the parameters are based on neighbors' behavior. The results revealed that a neighbor in foreclosure increases the hazard of defaults by 18 percent (Towe & Lawley, 2013).

Furthermore, a study was conducted to predict the borrower-related determinants that affect default risk in micro finance loans. The empirical results revealed that default risk is mainly linked to the borrower's high medical expenses, years of experience, education and additional liabilities with a high interest rate (Sandar, Charoenloet, & Sriwichailamphan, 2010).

The hypothesis in this study will be developed to examine the relationship between the borrower's demographic characteristics and the risk of default in housing loans as follows:

H_0: Insignificant relationship between borrower demographic characteristics represented by gender, marital status, existence of dependence, existence of additional collateral, nature of job occupation, job industry, job location, loan type and default risk.

H_1: Significant relationship between borrower demographic characteristics represented by gender, marital status, existence of dependence and additional collateral, nature of job occupation, job industry, job location, loan type, and default risk.

3. ECONOMETRICS MODEL

3.1. Model Selection Criteria

We will review similar studies that were conducted before and examine the factors driving default risk to decide on which model shall be adopted in this study. In 2013, a study was performed to assess the likelihood of default for US banks using a range of statistical methodologies, including linear discriminate analysis (LDA), probit model and logistic regression. The author examined a sample of 298 American commercial banks for model estimation collected between 2007 and 2010 during the financial crisis. The stepwise selection was applied for the logit and probit models. The logit and probit models have roughly similar explanatory capacities based on the fit in the training data while the LDA model had less explanatory power. The empirical results revealed that the logit model outperformed the probit and LDA models with an average fit of 80.4% compared to 62.2% and 42.6% respectively. According to the results of the ROC study, the logit model performed the best with an area under the curve (AUC) of 96.48% compared to the probit and LDA models which had 82.28% and 83.52% respectively. AUC offers a straightforward measure of merit for the effectiveness of the built-in classifier. Overall, the findings of these analyses supported the notion that the logit model performed better than other models when applied to both training data and test data (Gurný & Gurný, 2013).

Since default is a binary variable and loans are either defaulted or not defaulted, we will use binary logistic regression, a modeling technique used to predict probability for dependent variables that exist in default or normal (binary) form.

3.2. Binary Logistic Regression Model

The binary logistic regression analysis is used in this study to model the probability of the occurrence of one of the two possible outcomes which is the default or non-default based on one or more predictor variables such as borrower parameters. In binary logistic regression, the logistic function (also known as the sigmoid function) is used to model the relationship between predictor variables and the log-odds of the result. The logistic function guarantees that the expected probability falls between 0 and 1 which is appropriate for binary outcomes. The logistic regression model calculates the coefficients (log-odds ratios) for each predictor independent variable, reflecting the direction and intensity of the link with the result. These coefficients are usually computed using maximum likelihood estimation (Hosmer, Lemeshow, & Sturdivant, 2013).

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k \quad (1)$$

Where P is the probability of the outcome variable occurring.

$x_1, x_2, \ldots\ldots, x_k$ are the independent variables.

β_0 is the intercept.

$\beta_1, \beta_2, \ldots\ldots \beta_k$ are the coefficients to be estimated.

We apply the logistic function to transform the linear combination of predictor variables into a probability between 0 and 1.

$$P = \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k)}} \qquad (2)$$

Where e is the base of the natural logarithm.

The method used to estimate the coefficient $\beta_0, \beta_1, \beta_2, \ldots \beta_k$ is the maximum likelihood estimation method. This method pursues finding the set of coefficients that maximizes the likelihood of the given data used in the assumed logistic regression model in order to get the best fitting model parameters (Hosmer et al., 2013).

4. METHODOLOGY

4.1. Demographic Parameters

Most of the demographic parameters used in this study were extracted from the independent variables mentioned in the literature review and had a significant relationship with the default rates. In addition, the independent variables are referred to as per the below definitions and sub-categorical parameters:

- Gender: Male, female and couples co-borrower male and female.
- Marital Status: Single, married, widowed and divorced borrowers.
- Job Category: Employed, self-employed and freelancer.
- Job Economical Sector: Banking, service, public, private commercial, industrial and construction sectors.
- Job Location: Whether the borrower works inside Lebanon or expatriate.
- Existence of additional collateral: The presence of a personal guarantor, financial or asset collateral.
- Mortgage Location: It refers to the location of the apartment subject of the loan whether in Beirut, Mount Lebanon, South Lebanon, North Lebanon and West Lebanon.
- Requested housing loan purpose: Purchase, construction, under-construction or renovation.

4.2. Parameter Selection Criterion and Model Development

We first perform a single regression analysis for each explanatory variable then we save explanatory variables whose significance of the Wald test is associated with a p-value less than 0.25 and drop those whose variables are associated with a p-value greater than 0.25 to develop the best-fitted binary logistic regression model. The screening criterion for variable selection p-value thresholds of 0.25 is based on the recommendations of the study conducted by Bendel and Afifi (1977) and Mickey and Greenland (1989). Next, we will drop predictors for evidence of multicollinearity. Then, we run a multivariable regression analysis including all variables whose single p-values are less than 0.25. Next, we select variables that have p-value less than 0.1 and eliminate those with p-values greater than 0.1 as insignificant predictors. The new reduced model will be compared to the previous full model using the likelihood ratio test. Next, we refine the main effects model and ensure that selection variables are scaled correctly. The next procedure is to check if there are any interactions among the predicted variables. We add an interaction variable to a new model and compare it to the previous one using the likelihood ratio test. The interaction variable is created as the arithmetic product of the pairs of main effect explanatory variables. After obtaining the fitted model, we perform a diagnosis test (Neyman, 2023).

4.3. Model Diagnosis Test

The diagnosis procedures applied in this study include Hosmer-Lemeshow test for overall goodness of fit, likelihood ratio test, model adequacy test through link test, model accuracy and classification table, and Roc curve (Neyman, 2023).

Nurture: *Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN: *1994-1633/* ***Print ISSN***: *1994-1625*
DOI: *10.55951/nurture.v18i4.901*| ***URL***: www.nurture.org.pk
Publisher: *Nurture Publishing Group*

4.4. Stepwise Regression

We employ a stepwise technique to apply binary logistic regression analysis to the data set serving as a judge for the explanatory variable's relative relevance signifies attaining the optimal logistic model in the end.

4.5. Methodology Framework

The methodology framework of this study is displayed as below:

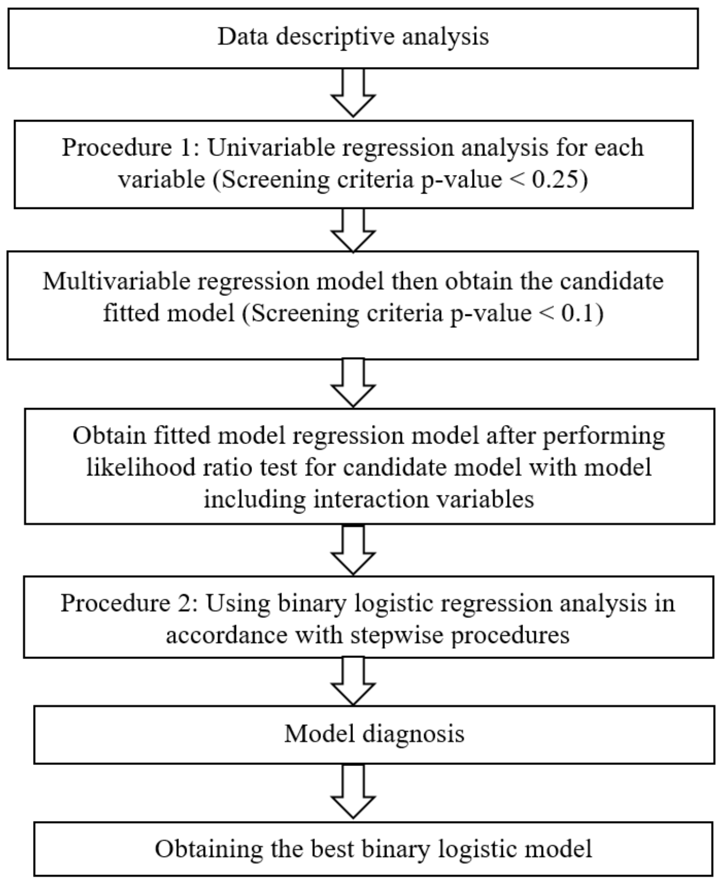

Figure 1. Methodology framework.

Source: Research finding.

Figure 1 illustrates the step-by-step methodology adopted in this study. First, we will perform a data analysis for individual borrowers followed by applied procedures to obtain the best-fitted model that describes the relationship between the log odd of default and the categorical borrower demographic predictors. The model then will be tested to ensure its accuracy and adequacy.

5. DATA ANALYSIS AND MODEL DEVELOPMENT

5.1. Descriptive Analysis

The application concerning the binary logistic regression analysis is applied to lending portfolio data using the Stata statistical program. The data were obtained from a leading bank in Lebanon that specializes in granting housing loans. The task is predicting whether a borrower would default or not for the loan he/she had granted. The borrower's data were labeled and categorized as 1 for defaulted borrowers and 0 for non-defaulting borrowers. The study used data of 6743 borrowers that are categorized as defaulted borrowers and counted 566 borrowers representing 8.4% of the total number of borrowers in the sample and 6177 borrowers are non-defaulting representing 91.6 % of the total sample size. Table 1 presents the details about the frequency and percentage distribution of the groups.

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

Table 1. Distribution of loan status.

Class name	Frequency	Distribution
Default	566	8.39%
Non-default	177	91.61%
Total	6743	100%

Source: Research finding.

According to Table 1, the number of borrowers who are defaulted is 566 representing 8.39% of the observations compared to 6177 borrowers who represent 91.61% of the observations that are non-default borrowers. The sample size of the data is 6742 observations. The statistical analysis information is displayed in Table 2 which also includes the explanatory factors' means and standard deviations.

Table 2. Descriptive statistical analysis.

Explanatory variables	Sub-category	Freq.	Percent	Cum.
X1 = Gender	Female	377	5.59	5.59
	Male	669	9.92	15.51
	MF	5,697	84.49	100
X2 = Marital status	Divorced	225	3.34	3.34
	Married	5,697	84.49	87.82
	Single	776	11.51	99.33
	Widow	45	0.67	100
X3 = Existing children	No	1,351	20.04	20.04
	Yes	5,392	79.96	100
X4 = Additional guarantee	No	6,255	92.76	92.76
	Yes	488	7.24	100
X5 = Job category	Employee	5,179	76.81	76.81
	Self-employed	1,564	23.19	100
X6 = Job industry	Banking	456	6.76	6.76
	Commercial	382	5.67	12.43
	Construction	72	1.07	13.5
	Industrial	74	1.1	14.59
	Public	605	8.97	23.57
	Service	5,154	76.43	100
X7 = Job location	Expatriate	1,349	20.01	20.01
	Resident	5,394	79.99	100
X8 = Mortgage location	Beirut	660	9.79	9.79
	Mount Lebanon	4,760	70.59	84.67
	North Lebanon	603	8.94	93.61
	South Lebanon	431	6.39	100
X9 = Loan type	Construction sector	430	6.38	6.38
	Purchase	6,183	91.7	98.07
	Renovation	65	0.96	99.04
	Under construction	65	0.96	100

Source: Research finding.

Data analysis revealed the following results:

- Regarding marital status, married borrowers represent 83.57 percent of defaulted clients while single borrowers represent 10.25 percent.
- 90.64 percent of defaulted borrowers have no additional collateral.
- Employee borrowers present 64.5 percent of defaulted borrowers.

Nurture: *Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN: *1994-1633/* ***Print ISSN***: *1994-1625*
DOI: *10.55951/nurture.v18i4.901*| ***URL***: www.nurture.org.pk
Publisher: *Nurture Publishing Group*

- Defaulted borrowers working in the service sector present 71.38 percent whereas those work in commercial sector present 10.95 percent.
- Borrowers who have jobs outside Lebanon present 13.25 percent of defaulted borrowers.
- Loans granted for purchase apartments in the Mount Lebanon area present 70 percent of defaulted loans.
- There are 89.22 percent of defaulted loans granted for purchase purposes.

5.2. Model Development

5.2.1. Fit Single Predictor Models: Keep Predictors Have Significance Levels Less Than 0.25 Using the Wald Z-Score

We run nine single logistic regressions and explain significant explanatory variables with a significance level less than 0.25.

Table 3. Significance of Wald for predicted variables summary table.

Predictors	**Significance of Wald Z**	**Remark**
Gender		
M	0.426	Drop
MF	0.814	Drop
Marital status		
Married	0.087	Consider further and the p-value is < 0.25.
Single	0.054	Consider further and the p-value is < 0.26.
Widow	0.129	Consider further and the p-value is < 0.27.
Existence of dependent	0.303	Drop
Additional guarantee		
Yes	0.042	Consider further and the p-value is < 0.25.
Job category		
Self-employed	0	Consider further and the p-value is < 0.25.
Job industry		
Commercial	0	Consider further and the p-value is < 0.25.
Construction	0.26	Consider further and the p-value is < 0.25.
Industrial	0	Consider further and the p-value is < 0.25.
Public	0.011	Consider further and the p-value is < 0.25.
Service	0.033	Consider further and the p-value is < 0.25.
Country		
Lebanon	0	Consider further and the p-value is < 0.25.
City		
West Lebanon	0.112	Consider further and the p-value is < 0.25.
Mount Lebanon	0.538	Consider further
North Lebanon	0.564	Consider further
South Lebanon	0.318	Consider further
Loan type		
Purchase	0.013	Consider further and the p-value is < 0.25.
Renovation	0.061	Consider further and the p-value is < 0.25.
Under-construction	0.185	Consider further and the p-value is < 0.25.

Source: Research finding.

Table 3 presents the outcome of single logistic regression and the significance level of predictors using the Wald Z-Score. The explanatory variables gender and existence of dependent have P value greater than 0.25 and therefore, they are not statistically significant and will be dropped. We will keep the sub-categorical parameters related to the mortgage location for further consideration.

5.2.2. Examine Potential Predictors for Multi-Collinearity Indicators

Collinearity refers to a situation in statistical models especially in regression analysis where two or more predictor variables (independent variables) are highly correlated with each other. When correlation occurs, it means that one of the predictor can be linearly predicted from the others. Variance Inflation Factor (VIF) is a measure used to

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

detect multi-collinearity in regression analysis. It quantifies how much the variance of an estimated regression coefficient is increased as a result of collinearity. Generally, a VIF value greater than 10 indicates the presence of multi-collinearity and that the associated independent variable is highly correlated with other independent variables in the model (Midi, Sarkar, & Rana, 2010).

Since explanatory variables are qualitative, no correlation exists among any of the predicted variables.

5.2.3. Run Multivariable Regression Including Predictors Associated with P-Value < 0.25

We run multiple predictor models using the explanatory variables including marital status, additional guarantee, job category, job industry, country, city and loan type.

Table 4. Full model with predictors of p-value < 0.25.

Logistic regression						
Number of observations	6,743					
LR chi2(18)	110.91					
Prob > chi2	0					
Log likelihood	-1888.457	Pseudo R2		0.0285		
Loan status	**Odds ratio**	**Std. err.**	**Z**	**P>z**	**[95% conf.**	**interval]**
Marital status						
Married	0.685	0.148	-1.74	0.082	0.448	1.049
Single	0.659	0.166	-1.65	0.099	0.402	1.081
Widow	1.887	0.818	1.47	0.143	0.806	4.415
Additional guarantee						
YES	1.232	0.191	1.34	0.18	0.908	1.671
Job category						
Self-employed	1.703	0.172	5.26	0	1.396	2.078
Job industry						
Commercial sector	2.519	0.667	3.49	0	1.498	4.235
Construction sector	1.149	0.560	0.29	0.775	0.441	2.988
Industrial sector	3.935	1.434	3.76	0	1.926	8.038
Public sector	1.774	0.456	2.23	0.026	1.072	2.937
Service sector	1.398	0.314	1.49	0.135	0.900	2.171
Country						
LEBANON	1.447	0.193	2.77	0.006	1.114	1.881
City						
Beqaa	1.319	0.325	1.12	0.261	0.813	2.141
Mount Lebanon	1.159	0.184	0.93	0.351	0.849	1.583
North Lebanon	1.157	0.246	0.69	0.493	0.762	1.756
South Lebanon	1.294	0.294	1.14	0.256	0.829	2.021
Loan type						
Purchase	0.793	0.131	-1.39	0.163	0.573	1.098
Renovation	0.266	0.196	-1.79	0.073	0.062	1.131
Under construction	1.636	0.610	1.32	0.187	0.787	3.401
Intercept	0.058	0.022	-7.19	0	0.026	0.126

Source: Research finding.

Table 4 presents the output of regression analysis and the significant level of predictions where the selection criteria of parameters are those who are associated with a p-value less than 0.25.

*Nurture: Volume 18, Issue 4, 847-867, **2024***
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

Table 5. Significance of Wald for predicted variables summary table P- value < 0.1

Predictors	Significance of Wald Z	Remark
	P>z	
Marital status		
Married	0.082	Consider further and the p-value is < 0.10.
Single	0.099	Consider further and the p-value is < 0.10.
Widow	0.143	Consider further
Additional guarantee		
YES	0.18	Drop
Job category		
Self-employed	0	Consider further and the p-value is < 0.10.
Job industry		
Commercial sector	0	Consider further and the p-value is < 0.10.
Construction sector	0.775	Consider further
Industrial sector	0	Consider further and the p-value is < 0.10.
Public sector	0.026	Consider further and the p-value is < 0.10
Service sector	0.135	Consider further
Country		
LEBANON	0.006	Consider further and the p-value is < 0.10.
City		
Beqaa	0.261	Drop
Mount Lebanon	0.351	Drop
North Lebanon	0.493	Drop
South Lebanon	0.256	Drop
Loan type		
Purchase	0.163	Consider further
Renovation	0.073	Consider further and the p-value is < 0.10.
Under construction	0.187	Consider further

Source: Research finding.

Table 5 presents the significance levels of predictors using the Wald Z-Score. The selection criteria of significant parameters are set to those parameters that are associated with a p-value level of significant less than 0.1.
The explanatory variables additional guarantee and city have P values greater than 0.1 and therefore they are not statistically significant and will be dropped. We will keep the sub-categorical parameters related to marital status, job industry and loan type for further consideration.

5.2.4. Multivariable Regression with Predictors of p-Values < 0.10

We run a multivariable regression model that includes predictors associated with p-values < 0.10.

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

Table 6. Multivariable model.

Logistic regression						
Number of observations	6,743					
LR chi2(13)	107.39					
Prob > chi2	0					
Log likelihood	-1890.2177	Pseudo R2	0.0276			
Loan status	**Odds ratio**	**Std. err.**	**Z**	**P>z**	**[95% conf.**	**interval]**
Marital status						
Married	0.698	0.149	-1.7	0.089	0.452	1.057
Single	0.66	0.167	-1.62	0.105	0.405	1.088
Widow	1.89	0.822	1.48	0.139	0.812	4.439
Job category						
Self-employed	1.710	0.172	5.31	0	1.40	2.084
Job industry						
Commercial sector	2.638	0.695	3.68	0	1.573	4.422
Construction sector	1.154	0.562	0.29	0.768	0.444	3.00
Industrial sector	4.022	1.463	3.83	0	1.971	8.206
Public sector	1.800	0.461	2.29	0.022	1.088	2.976
Service sector	1.420	0.318	1.56	0.118	0.915	2.204
Country						
LEBANON	1.427	0.190	2.67	0.008	1.099	1.853
Loan type						
Purchase	0.770	0.125	-1.6	0.109	0.560	1.060
Renovation	0.258	0.190	-1.83	0.067	0.061	1.097
Under construction	1.561	0.577	1.2	0.228	0.756	3.224
Intercept	0.069	0.025	-7.29	0	0.033	0.142

Source: Research finding.

Table 6 presents the output of multivariable regression analysis and the significant level of selected predictions where the selection criteria of parameters are those associated with a p-value less than 0.1.

5.2.5. Compare Models Obtained from Previous Steps Using a Likelihood Ratio Test

Model 3: (-2) ln L = (-2) X (-1888.4574) = 3777 Deviance d f = 6743 - (18) = 6725

Model 4: (-2) ln L = (-2) X (-1890.2177) = 3780 Deviance d f = 6743 - (13) = 6730

5.2.5.1. (LR) Test Manual Calculation

LR Test = [(-2) ln (L_r)] - [(-2) ln (L_f)] == 3780- 5777 =3.52.

LR Test d f = Change deviance d f = Change in numbers predictors in model = 6730 - 6725= 5.

P-value = Pr {Chi square with 5 degrees of freedom > 3.42} = 0.62.

This is not significant. We drop additional guarantees and city explanatory variables.

5.2.5.2. Likelihood Ratio Test Using Stata Software

Running the quietly procedure on the full and reduced model, we perform the likelihood ratio test and the stata results are as follows: LR chi2(5) = 3.52 Prob > chi2 = 0.6203 Matched.

5.2.6. Investigate Confounding Predictors

The good-fitted final model is the five predictors' model, i.e., marital status, job category, job industry, country and loan type. Next, we will assess these variables as potential confounders using the following two criteria:

1. Predictors with p-value < 0.10.

Online ISSN: 1994-1633/ ***Print ISSN***: 1994-1625
DOI: 10.55951/nurture.v18i4.901| ***URL***: www.nurture.org.pk
Publisher: Nurture Publishing Group

2. Relative Change in estimated betas > 15% using the following formula (Neyman, 2023):

$$\Delta\hat{\beta} = \frac{|\hat{\beta}\text{without confounders} - \hat{\beta}\text{with confounders}|}{\hat{\beta}\text{with confounders}} \text{ x } 100 \qquad (3)$$

We need to find the change in estimated betas for both reduced and full model and compare it to the threshold of 15%. A summary table after finding the result through Stata software is below.

Table 7. Investigate confounding.

Predictors	**Coefficient reduced model**	**Coefficient full model**	**Change in expected betas**	**% change**
Marital status				
Married	-0.368	-0.377	0.008	-2.34%
Single	-0.408	-0.415	0.007	-1.71%
Widow	0.641	0.635	0.006	0.96%
Job category				
Self-employed	0.536	0.532	0.003	0.74%
Job industry				
Commercial	0.970	0.924	0.046	4.98%
Construction	0.143	0.139	0.004	3.16%
Industrial	1.391	1.370	0.021	1.59%
Public	0.587	0.573	0.014	2.47%
Service	0.350	0.335	0.015	4.61%
Country				
Resident	0.356	0.370	-0.014	-3.80%
Loan type				
Purchase	-0.260	-0.230	-0.029	12.88%
Renovation	-1.35	-1.322	-0.029	2.23%
Under construction	0.445	0.492	-0.047	-9.54%

Source: Research finding.

Table 7 presents the relative change of estimated betas obtained from the reduced and full models. The value is compared to the threshold of 15 percent. The relative change in the betas in the good model is less than 15 %.

5.2.7. Examine Effect Modification

Effect modification is needed to be investigated to check if the relationship between the predictor and the dependent variable changes depending on the interaction between the predictors with another predictor in the model that is whether the effect of one variable on the outcome is modified by the presence or level of another variable. Effect modification will be detected by including the interaction terms in the fitted candidate logistic regression model and assessing its significance (Hosmer et al., 2013).

Table 8. The regression model includes interaction variables.

Jobcode _ industrycode interaction variable	**Freq.**	**Percent**	**Valid**	**Cum.**
Employee and banking sector	453	6.72	6.72	6.72
Employee and commercial sectors	187	2.77	2.77	9.49
Employee and construction sectors	38	0.56	0.56	10.05
Self-employee and commercial sectors	251	3.72	3.72	13.78
Employee and public sectors	587	8.7	8.71	22.48
Employee and service sectors	3898	57.8	57.81	80.29
Self-employed and industrial sectors	21	0.31	0.31	80.6
Self-employed and public sectors	18	0.27	0.27	80.87
Self-employee and service sectors	1290	19.13	19.13	100
Total	6743	99.99	100	200

Source: Research finding.

Nurture: *Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN: *1994-1633/* ***Print ISSN***: *1994-1625*
DOI: *10.55951/nurture.v18i4.901|* ***URL***: www.nurture.org.pk
Publisher: *Nurture Publishing Group*

Are borrowers who are both self-employed and who work in the banking sector more likely to default? For this scenario, we will develop a new interaction variable titled Jobcode_ Industrycode to deduct borrowers whose job is in the banking industry and are employees. Next, we run a regression model that includes the main effects of both of the variables contributing to the interaction (Neyman, 2023).

Table 8 presents statistical analysis that reflects all the possible outcomes of the interaction between job category and job industry parameters.

From the table above, 453 borrowers are employed in the banking sector.

Running a logistic regression using the Stata software includes the predicted variables of the final model in addition to the interaction variable.

Full Model: It predicted variables, marital status, job category, job industry, country loan type Jobcode_Industrycode.

Table 9. The regression model includes interaction explanatory variables.

Logistic regression						
No. of observations	6,743					
LR chi2(14)	107.84					
Prob > chi2	0					
Log likelihood	-1889.990	Pseudo R2	0.027			
Loan status	**Odds ratio**	**Std. err.**	**z**	**P>z**	**[95% conf.**	**interval]**
Marital status						
Married	0.693	0.150	-1.68	0.092	0.453	1.061
Single	0.665	0.167	-1.62	0.106	0.405	1.090
Widow	1.918	0.831	1.5	0.133	0.820	4.486
Job code						
Self-employed	2.238	0.925	1.95	0.051	0.995	5.034
Industry code						
Commercial sector	2.499	0.693	3.3	0.001	1.451	4.305
Construction sector	1.184	0.580	0.35	0.729	0.453	3.095
Industrial sector	4.571	1.878	3.7	0	2.042	10.230
Public sector	2.200	0.867	2	0.045	1.016	4.765
Service sector	1.847	0.833	1.36	0.174	0.762	4.472
Country						
LEBANON	1.434	0.191	2.71	0.007	1.104	1.863
Loan type						
Purchase	0.772	0.125	-1.59	0.112	0.561	1.062
Renovation	0.261	0.192	-1.82	0.068	0.061	1.107
Under construction	1.571	0.581	1.22	0.222	0.760	3.24
Jobcode_industrycode	0.951	0.070	-0.67	0.502	0.822	1.100
Intercept	0.071	0.026	-7.1	0	0.034	0.148

Source: Research finding.

Table 9 presents the output of the regression analysis that includes the significant predictors and the interaction variable of the full model.

Next, run a logistic regression using the Stata software that includes the predicted variables of the final model, this is the reduced model.

Reduced Model: It includes predicted variables, marital status, job category, job industry, country and loan type.

Nurture: *Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN: *1994-1633/* ***Print ISSN***: *1994-1625*
DOI: *10.55951/nurture.v18i4.901*| ***URL***: www.nurture.org.pk
Publisher: *Nurture Publishing Group*

Table 10. Reduced model.

Logistic regression						
No. of observations	6,743					
LR chi2(13)	107.39					
Prob > chi2	0					
Log likelihood	-1890.2177	PseudoR2	0.0276			
Loan status	**Coefficient**	**Std. err.**	**z**	**P>z**	**[95% conf.**	**interval]**
Marital status						
Married	-0.368	0.216	-1.7	0.089	-0.793	0.056
Single	-0.408	0.251	-1.62	0.105	-0.902	0.085
Widow	0.641	0.433	1.48	0.139	-0.207	1.490
Job category						
Self-employed	0.536	0.101	5.31	0	0.338	0.734
Economical sector						
Commercial sector	0.970	0.263	3.68	0	0.453	1.486
Construction sector	0.143	0.487	0.29	0.768	-0.811	1.098
Industrial sector	1.391	0.363	3.83	0	0.678	2.104
Public sector	0.587	0.256	2.29	0.022	0.084	1.090
Service sector	0.350	0.224	1.56	0.118	-0.088	0.790
Country						
LEBANON	0.356	0.133	2.67	0.008	0.094	0.617
Loan type						
Purchase	-0.260	0.162	-1.6	0.109	-0.579	0.058
Renovation	-1.351	0.736	-1.83	0.067	-2.795	0.092
Under construction	0.445	0.369	1.2	0.228	-0.279	1.170
Intercept	-2.669	0.366	-7.29	0	-3.386	-1.951

Source: Research finding.

Table 10 presents the output of the regression analysis and the significant predictors that produce the best fit model.

Next, compare reduced and full models using a likelihood ratio test.
Full Model: (-2) ln L = (-2) X (-1889.990) = 3779.98 deviance d f = 6743 - (14) = 6729.
Reduced Model: (-2) ln L = (-2) X (-1890.217) = 3780.43 deviance d f = 6743 - (13) = 6730.

5.2.7.1. LR Test Comparing Previous Models by Manual Calculation
LR Test = [(-2) ln (L_r)] - [(-2) ln (L_f)] = 3780- 3779.98 = 0.4554.
LR Test d f = Change deviance d f = Change in numbers predictors in model = 6730 - 6729= 1.
P-value = Pr {chi square with 5 degrees of freedom > 2} = 0.50.
This is not significant. We drop additional guarantees and city explanatory variables.

5.2.7.2. LR Test Using Stata Software
Assumption: reduced nested within full LR chi2 (1) = 0.45 and Prob > chi2 = 0.5001 matched.
Therefore, we drop the interaction variable and the fitted model is the reduced one.
Therefore, a reasonable multiple predictor model of the outcome default in this sample includes the following predictors: Marital status, job category, job industry, country and loan type.

5.3. Final Model
Logit (Pr [default=1]) = -2.67- 0.37 X Married Borrower -0.41X Single + 0.64 X Widow + 0.53 X Self-employed +0.97 X Job in Commercial Sector +0.14 X Construction +1.39 X Industrial +0.58 X Public + 0.35 X Service Sector + 0.35 X Local

Nurture: *Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN: *1994-1633/* ***Print ISSN***: *1994-1625*
DOI: *10.55951/nurture.v18i4.901*| ***URL***: www.nurture.org.pk
Publisher: *Nurture Publishing Group*

Job -0.26 Purchase Loan – 1.35 X Renovation Loan + 0.44 X Under-Construction (4)

5.4. Stepwise Regression

Applying stepwise regression procedure using stata statistic software with level of significant P value less than 0.1 is reflected in the table.

Table 11. Stepwise regression.

Logistic regression						
Number of observations	6,743					
LR chi2(8)	100.17					
Prob > chi2	0					
Log likelihood	-1893.824	Pseudo R2	0.0258			
Loan status	**Coefficient**	**Std. err.**	**z**	**P>z**	**[95% conf.**	**interval]**
Industry code						
Commercial sector	0.652	0.151	4.31	0	0.35	0.948
Industrial sector	1.066	0.296	3.6	0	0.486	1.647
Public sector	0.271	0.154	1.76	0.078	-0.030	0.574
Loan type						
Purchase	-0.316	0.148	-2.13	0.033	-0.607	-0.025
Renovation	-1.407	0.733	-1.92	0.055	-2.845	0.029
Marital status						
Widow	0.976	0.379	2.57	0.01	0.231	1.720
Country						
LEBANON	0.335	0.132	2.54	0.011	0.076	0.595
Job code						
Self-employed	0.563	0.099	5.67	0	0.368	0.757
Intercept	-2.637	0.185	-14.25	0	-3.00	-2.274

Source: Research finding.

Table 11 presents the output of the regression analysis with the significant predictors that produce the best fit model using the stepwise regression procedures.

We can conclude that predictors that have a significant relationship with the risk of default that were obtained by selecting the best fitted regression model based on criteria adopted before and using the model of lowest AUC are the same ones obtained by using the stepwise regression procedures.

6. MODEL DIAGNOSIS

6.1. The Hosmer-Lemeshow Test of Goodness-of-Fit

Test of Hosmer-Lemeshow Goodness-of-Fit Hypothesis:

HO: The data and the existing model fit each other well. HA: not.

When the p-value is greater than 0.05, then the model fits the data reasonably well and failure to reject the null hypothesis occurs. However, when the p-value is less than or equal to 0.05, reject the null hypothesis and the model does not fit the data well. We use stata software command estat gof to obtain the results of Hosmer Lemeshow Test.

Nurture: *Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN: *1994-1633/* ***Print ISSN***: *1994-1625*
DOI: *10.55951/nurture.v18i4.901|* ***URL***: www.nurture.org.pk
Publisher: *Nurture Publishing Group*

Table 12. Goodness-of-fit test.

Number of observations = 6,743	Number of groups = 7
Hosmer Lemeshow chi2 (5) = 1.90	Prob > chi2 = 0.862

Table 12 presents the model diagnosis test using goodness of fit procedures.
The Hosmer_Lemeshow test (p=0.8625) suggests that the null hypothesis of "good fit" is not rejected.

6.2. The Link Test

A basic inspection of the fitted model is the link test. It evaluates whether the candidate-fitted model fits the data well enough (null hypothesis) or if it does not, if more modeling is necessary (alternative hypothesis)
H0: The data are suitably fitted by the existing model.
H1: We require other model.
A "null hypothesis" adequate model (reduced) is compared to an alternative hypothesis enhanced (full) model using the Likelihood Ratio (LR) Test.
Reduced Model: logit[p] = β (0) + β (1) [$\hat{p}$ model].
Full Model: logit[p] = β (0) + β (1) [$\hat{p}$ model] + β (2) [$\widehat{p\ 2}$ model].
H0: β2 = 0.
H 1 = not.
_hat: This is the predicted probability from candidate fitted model.
_hatsq: If the null is true (the model is adequate), this should be non-significant.
We expect the p-value for (_HAT) to be significant and the evidence of a good fit is reflected in a non-significant (_HATSQ) (Neyman, 2023).

Table 13. Likelihood test.

Iteration 0: Log-likelihood	-1943.9				
Iteration 1: Log-likelihood	-1902.8				
Iteration 2: Log-likelihood	-1890.2				
Iteration 3: Log-likelihood	-1890.2				
Iteration 4: Log-likelihood	-1890.2				
No. of observation	6743				
LR chi2(2)	107.48				
Prob > chi2	0				
Log likelihood	-1890.170	Pseudo R2	0.027		
Loan Status	Coefficient	Std. err.	z	P>z	[95%
_hat	1.190	0.621	1.91	0.056	-0.028
_hats	0.044	0.142	0.31	0.757	-0.234
Intercept	0.194	0.666	0.29	0.77	-1.112

Source: Research finding.

Table 13 presents the diagnosis test that checks whether the candidate-fitted model fits the data well enough.
The p-value of (_hat) is 0.056 which means it is significant. The p-value of (_hatsq) is equal to 0.757 which means it is not significant. Consequently, the null hypothesis of "model adequacy" is not rejected.

6.3. Classification Table

After checking the goodness-of-fitted model, we need to ensure that the individual predictions used in the fit model are correct most of the time. The classification table shows the difference between the expected and observed numbers of successes. Similarly, it compares the observed number of failures to the anticipated number of failures. Stata software by default chooses a threshold probability for an event as 0.5. This probability can be amended when needed.

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

Table 14. Classification table.

Description	True		Total
Classified	D	~D	
+	0	0	0
-	566	6177	6743
Total	566	6177	6743
Classified + if predicted Pr(D) >= 0.5			
True D defined as loan status != 0			
Sensitivity		Pr (+ D)	0.00%
Specificity		Pr (-~D)	100.00%
Positive predictive value		Pr (D +)	.%
Negative predictive value		Pr (~D -)	91.61%
False + rate for true ~D		Pr (+~D)	0.00%
False - rate for true D		Pr (- D)	100.00%
False + rate for classified		Pr (~D +)	.%
False - rate for classified		Pr (D -)	8.39%
Correctly classified		**91.61%**	

Source: Research finding.

Table 14 presents the classification table that shows the percentage of data used in the study correctly classified by the best fit model.

The output revealed that 91.61 percent of the data is correctly classified by the model.

6.4. The ROC Curve

After checking the goodness-of-fitted model by using the classification table, a graphical representation showcasing the overall effectiveness of a logistic model and the corresponding equation for predicted probabilities will be performed using the receiver operating characteristics curve (ROC curve).

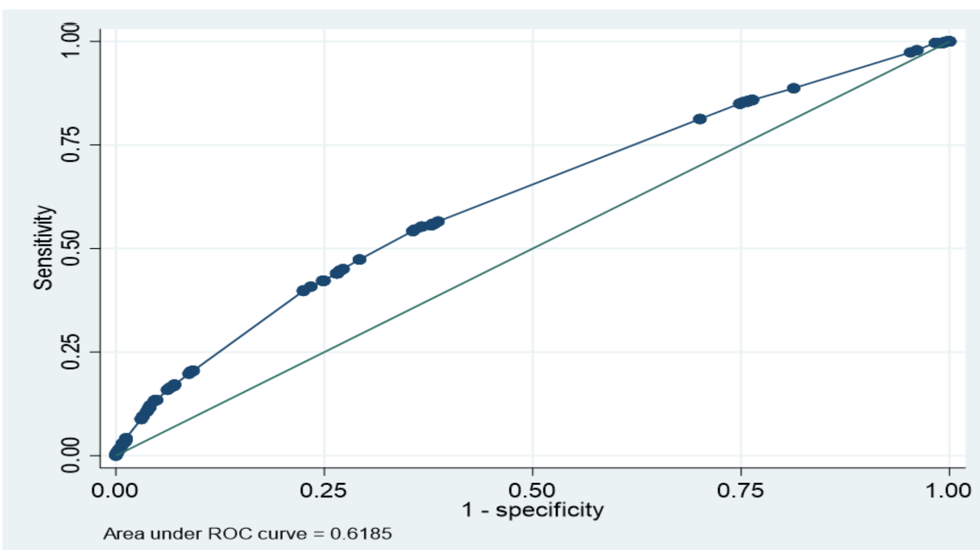

Figure 2. Roc curve.

Source: Research finding.

Figure 2 illustrates the overall effectiveness of a logistic model. The higher the value of the area under the curve (AUC), the higher is the effectiveness of the best fit model.

Nurture: Volume 18, Issue 4, 847-867, 2024
Online ISSN: 1994-1633/ Print ISSN: 1994-1625
DOI: 10.55951/nurture.v18i4.901| URL: www.nurture.org.pk
Publisher: Nurture Publishing Group

The reference line is the straight line with slope =1, it resembles the ROC curve where chance alone is effective (coin toss with probability heads =0.50). ROC c-statistic = 0.6185 means that the overall percentage of cases in the data set that are correctly classified is 61.85 percent.

7. RESULTS AND DISCUSSION

The empirical results suggest that giving the value of the intercept at -2.67 means that when all predictor variables are set to their lowest values, the estimated probability of default is approximately 6.5%. In addition, for the marital status predictor, having the divorced category as a reference variable, the log odds of default for single and married borrowers are lower compared to divorced borrowers by about 34 and 31 percent respectively. Married borrowers have a lower risk of default since they benefit from a dual income which enhances their financial stability and the ability to pay their monthly obligations. In addition, single borrowers have a lower default risk than divorced borrowers since they do not experience the financial burdens that often accompany divorce such as legal fees child support payments, etc. However, a widow borrower with a coefficient of (+0.64) indicates that widow borrowers have higher log odds of default risk compared with divorced borrowers. The odds ratio of 1.90 indicates that the odds of default for widow borrowers are about 90 percent higher than for divorced borrowers because widows often lose a primary or secondary source of income upon the death of the spouse. This will lead to a significant reduction in income which can severely impact their ability to meet their financial obligations. On the other hand, a divorced borrower is often involved in predictable financial change and might have ongoing support from an ex-spouse.

Second, for the job category predictor, having the employed borrower category as the reference variable, self-employed borrowers with a coefficient of 0.43 associated with an odds ratio equivalent to 1.54 indicates that the odds of default for self-employed borrowers are about 54 percent higher than employed borrowers. This is because self-employed borrowers experience more volatile and unpredictable income inflow compared to employees who earn regular and fixed salaries. This volatility makes it difficult to manage the monthly obligations and therefore increases the risk of default.

Third, for job industry predictors, having the banking sector as a reference variable, the coefficients 0.97, 0.14, 1.39, 0.58 and 0.35 are associated with odds ratios of 2.64, 1.15, 4.01, 1.79, and 1.42 for borrowers working in commercial, construction, industrial, public and private sectors respectively indicates that the log odds of default for borrowers working in commercial, construction, industrial, public and private sectors are higher than those working in banking sector by 164%, 15%, 300%, 79%, and 42% respectively compared to borrowers working in banking sector because jobs in Lebanese banking sector are more stable and secure earning increases on a yearly basis in addition to additional compensation and bonuses. Furthermore, the monthly payment of housing loans for borrowers working in the banking sector is systematically deducted from the employee salary account. However, borrowers working in other economic sectors are exposed to high volatility since they are sensitive to the Lebanese economic conditions and the customers' spending. During the economic crisis and high inflation, purchasing power will decrease and therefore, customer spending will decrease and affect negatively the commercial sector. This will lead to a decrease in salaries and increase the risk of default. Borrowers who work in the construction sector have a higher risk of default than those working in the banking sector but the lowest among borrowers working in the commercial, industrial, public and private sectors because the construction sector can benefit from large-scale projects funded by private investments which may continue even during the economic crisis. However, the results are opposite to the ones mentioned in the study conducted by scholars such as Banasik and Crook (2005) where authors found that construction and retail industries are volatile to any economic distress. Furthermore, the high probability of default for borrowers working in the industrial sector is because industrial sector depends on raw materials imported from countries outside Lebanon that are priced in foreign currencies. The large devaluation of the local Lebanese currency combined with the bank's capital control procedures reduces access to foreign currencies and therefore decreases the import activities and thus the industrial business activities. This leads to a negative impact on borrowers working in the industrial sector and increases their ability to not pay their monthly obligations. In addition, borrowers working in the public sector are paid their salary in the Lebanese pound due to currency devaluation. It is expected that default risk will be increased. However, since the public sector does not depend on consumer spending or imported raw materials, the default rate of borrowers working in the public sector is lower than those working in both industrial and commercial sectors in comparison

*Nurture: Volume 18, Issue 4, 847-867, **2024***
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

with banking sector employee borrowers. Private sector employees will have a higher default rate compared to banking sector employees but lower than industrial and commercial sectors since demand for essential products and services will continue even during the economic downturns. Therefore, some business areas in the private sector will be affected and volatile to economic conditions and others will not. In addition, many private companies in Lebanon pay total or partial salaries in foreign currencies and this enhances the borrower's ability to settle their housing loan monthly payments.

Fourth, for borrower job location, having a foreign country as the reference variable, the coefficient of working in the local country variable is 0.35 associated with an odd ratio of 1.42 which indicates that borrowers working locally compared to those working in foreign countries have higher odds of default by 42% because borrowers having a job outside Lebanon will earn an income salary in foreign currency and therefore borrower income will be independent of any Lebanese economic instability. Borrowers who have a local job will be highly affected especially if he or she earns the income salary in Lebanese pounds and therefore, their ability to pay the loan monthly payment will be decreased.

Fifth, for the type of housing loan independent predictor, having construction loans as the reference variable, the log odds of purchase, renovation and under-construction loan variables are 0.77, 0.26 and 1.55 respectively. The empirical results suggest that borrowers who are granted purchase loans have lower log odds of defaulting on a loan compared to borrowers who took construction loans. The odds ratio of 0.77 indicates that the odds of default for purchase loan borrowers are about 23% lower than the construction loan borrowers. Similarly, borrowers who are granted renovation loans have lower log odds of defaulting on a loan compared to borrowers who took construction loans. The odds ratio of 0.26 indicates that the odds of default for renovation loan borrowers are about 75% lower than the construction loan borrowers because borrowers who request a renovation loan already own the housing unit subject to renovation. Renovation loans can increase the borrower's home equity and this may reduce the risk of default. However, borrowers who are granted under-construction loans have higher log odds of defaulting on a loan compared to borrowers who took construction loans. The odds ratio of 1.55 indicates that the odds of default for under-construction loan borrowers are about 55% higher than the construction loan borrowers. Borrowers who request under-construction projects are exposed to factors that increase the likelihood of loan default since these projects are subject to various risks including construction delays, cost overruns, contractors, and regulatory issues. These uncertainties can erode the borrower's finances leading to a higher risk of default. In addition, during the period needed to finalize the construction work, the market price of construction materials may change. For instance, the prices of construction materials highly increase when the Lebanese economy suffers from high inflation rates and due to the Lebanese pound devaluation. This leads to an increase in the default risk rates for borrowers requesting under-construction loans.

8. CONCLUSION

In this research, the researcher examines the relationship between factors related to the borrower's demographic characteristics and their influence on the risk of default in residential mortgage loans. The binary logistic regression analysis technique is applied to real data extracted from Lebanese financial institutions' housing lending portfolios to predict the impact of demographic characteristics on mortgage default risk. Two procedures have been used. The first method relied on the univariate Wald statistic and the likelihood ratio test for the significant coefficient in multivariable regression. The second method is adopted based on the stepwise regression technique. The analysis demonstrated that the two procedures selected the same predictors that fitted the optimal model which include the borrower's marital status, the borrower's job location and nature of job occupation, the economic sector and the type of requested housing loan. In addition, the applied regression procedures eliminate the remaining predictors which stood for gender, the existence of dependent, the existence of additional collateral, and mortgage location. These factors were excluded from the analysis since they were unable to provide a beneficial impact.

In addition, findings of the analysis also showed that the performance of the binary logistic regression analysis demonstrates the overall percentages who are correctly classified are 91.61%. The optimal model is determined by examining the model's overall goodness-of-fit and the parameter values and their signs within the binary logistic regression equation numbered (5) provided earlier.

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

The limitation of the study is reflected by the lack of information about borrowers' level of education and the years of work experience they have since these two variables might have a significant influence on the risk of default. In addition, there is also a lack of information about the borrower's age and the existence of loans other than housing loans or any other type of liabilities that affect the borrower's capabilities to pay back the loan.

FUNDING
This study received no specific financial support.

INSTITUTIONAL REVIEW BOARD STATEMENT
The Ethical Committee of the Beirut Arab University, Lebanon has granted approval for this study.

TRANSPARENCY
The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

COMPETING INTERESTS
The authors declare that they have no competing interests.

ARTICLE HISTORY
Received: 22 May 2024/ Revised: 16 September 2024/ Accepted: 27 September 2024/ Published: 11 October 2024

REFERENCES

Banasik, J., & Crook, J. (2005). *Explaining aggregate consumer delinquency behaviour over time*. Retrieved from Credit Research Centre, University of Edinburgh, Working Paper Series, No. (05/03):

Bendel, R. B., & Afifi, A. A. (1977). Comparison of stopping rules in forward "stepwise" regression. *Journal of the American Statistical Association*, *72*(357), 46–53. https://doi.org/10.2307/2286904

Borzekowski, R., & Cohen-Cole, E. (2008). *The effect of mortgage payment resets on default rates*. Retrieved from Federal Reserve Bank of Boston Working Paper Series, No. 08-5:

Fisher, L., & Gervais, S. (2011). Why do homeowners default? A study of self-reported reasons for default and their effect on the likelihood of default. *Journal of Housing Economics*, *20*(4), 302-316. https://doi.org/10.1016/j.jhe.2011.07.004

Foote, C. L., Gerardi, K., & Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics*, *64*(2), 234-245. https://doi.org/10.1016/j.jue.2008.07.006

Goel, A., & Rastogi, S. (2023). Understanding the impact of borrowers' behavioural and psychological traits on credit default: Review and conceptual model. *Review of Behavioral Finance*, *15*(2), 205-223. https://doi.org/10.1108/rbf-03-2021-0051

Goodman, L., S., Seidman, E., & Zhu, J. (2013). The case for principal reductions: How to improve the effectiveness of the home affordable modification program. *Journal of Structured Finance*, *19*(1), 50-61.

Gurný, P., & Gurný, M. (2013). Comparison of credit scoring models on probability of default estimation for us banks. *Prague Economic Papers*, *22*(2), 163-181.

Hosmer, J. D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.): John Wiley & Sons. https://doi.org/10.1002/9781118548387

Jimenez, G., Salas, V., & Saurina, J. (2006). Determinants of collateral. *Journal of Financial Economics*, *81*(2), 255-281. https://doi.org/10.1016/j.jfineco.2005.06.003

Lee, J. W., & Sohn, S. Y. (2021). Evaluating borrowers' default risk with a spatial probit model reflecting the distance in their relational network. *PloS One*, *16*(12), e0261737. https://doi.org/10.1371/journal.pone.0261737

Mickey, R. M., & Greenland, S. (1989). The impact of confounder selection criteria on effect estimation. *American Journal of Epidemiology*, *129*(1), 125-137. https://doi.org/10.1093/oxfordjournals.aje.a115101

Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, *13*(3), 253-267.

Munnell, A. H., Tootell, G. M., Browne, L. E., & McEneaney, J. (1996). Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review*, *86*(1), 25-53.

Nurture**: Volume 18, Issue 4, 847-867, **2024
***Online ISSN**: 1994-1633/ **Print ISSN**: 1994-1625*
***DOI**: 10.55951/nurture.v18i4.901| **URL**: www.nurture.org.pk*
***Publisher:** Nurture Publishing Group*

Neyman, J. (2023). *Unit 7: Logistic regression*. Retrieved from https://people.umass.edu/biep640w/pdf/7.%20%20Logistic%20Regression%202023.pdf

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, *8*(11), 169. https://doi.org/10.3390/data8110169

Sandar, T. M., Charoenloet, V., & Sriwichailamphan, T. (2010). Factor affecting the borrowers' repayment performance of microfinance in sub-urban areas of Yangon, Myanmar. Doctoral Dissertation, Graduate School, Chiang Mai University.

Towe, C., & Lawley, C. (2013). The contagion effect of neighboring foreclosures. *American Economic Journal: Economic Policy*, *5*(2), 313-335. https://doi.org/10.1257/pol.5.2.313

Nurture*: Volume 18, Issue 4, 847-867,* ***2024***
Online ISSN*: 1994-1633/* ***Print ISSN****: 1994-1625*
DOI*: 10.55951/nurture.v18i4.901|* ***URL****:* www.nurture.org.pk
Publisher: *Nurture Publishing Group*