THE UNIVERSITY *of* EDINBURGH

Novel Information in Estimating Loss Given Default in Brazil

Angela Rita Freitas De Moraes

Doctor of Philosophy

The University of Edinburgh

2018

Abstract

The Basel Accord regulates risk and capital requirements to ensure that a bank holds capital proportional to the exposed risk of its lending practices. Basel II allows banks to develop their own empirical models based on historical data for probability of default (PD), loss given default (LGD) and exposure at default (EAD). Brazil was among the first emerging market countries to release a timetable for the implementation of the Basel II Accord and aimed to apply it uniformly to all Brazilian financial institutions from 2005 to 2011. Within this context, the necessity arises of conducting research that could assist the financial institutions in improving the accuracy of their models.

This thesis has three objectives. The first is to develop a macro-economic model to predict the behaviour of the aggregate delinquency in Brazilian consumer loans. The model consists in testing co-integrating relationships and then estimating a short run error correction model. The results based on monthly data from 2000 to 2012 show that the delinquency rate is particularly sensitive to shocks on GDP and to the variation of workers' income. The analysis then shifts to micro or account level to model the behaviour of borrowers and certain novel types of information that can be used for prediction.

Second, customers fail to make loan repayments for a number of reasons, ranging from simple forgetfulness to deliberate attempts. For this reason, the second objective is to investigate the reasons for default and to explore ways of incorporating these variables into Recovery Rate (RR = 1 - LGD) models, since the standard approach overlooks real reasons for default and uses proxies for them such as marital status and length of employment. Customers who failed to repay their loans were interviewed in order to discover the causes for this failure. In addition, the interviews included questions aimed to measure the customer's personality traits and their financial knowledge in relation to the reasons for default. The empirical results show that the variables proposed in this study, namely, reason for missing payment, financial knowledge and risk taken, improve the prediction of the recovery rate.

Thirdly, it is known that recovery depends on the debt collection process and on the different options or actions that collection departments can take. Yet there is practically no literature exploring the impact of the lender's collection actions on RR/LGD. This work fills this gap by investigating the role of different collection actions at the loan-level for a retail credit product, and by estimating LGD models using Panel Data regressions.

Statement of original authorship

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Angela Rita Freitas De Moraes

October, 2018

Dedicated to my family

For their love, endless support

And encouragement

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisors Dra Galina Andreeva and Prof. Jonathan Crook for the continuous support of my research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I am very indebted to the Credit Research Centre for funding the survey which was part of this research. I am deeply grateful to the company that provided the data for this analysis

To my friends, thank you for listening, offering me advice, and supporting me through this entire process. Special thanks to Joanna and Mona whose companionship has made this process less painful.

Last but not least, I would like to thank my family: my daughter, my brothers and sisters for supporting me spiritually throughout this process and Fernando for being understanding and standing by my side.

.

Thank you.

Angela Freitas De Moraes

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADF – Augmented Dickey Fuller
AIRB – Advanced Internal Ratings-Based
ARMA – Auto Regressive Moving Average
BCBS – Basel Committee on Banking Supervision
BIS – Bank for International Settlements
CCI – Consumer Confidence Index
DSR – Debt Service Ratio
EAD – Exposure at default
ECL – Expected Credit Loss
FICO – Fair Isaac Corporation
GDP – Gross Domestic Product
IAS – International Accounting Standard
IASB – International Accounting Standards Board
IBGE – Brazilian Institute of Geography and Statistics
IFRS – International Financial Reporting Standard
IRB – Internal Ratings-Based
KMV – Kealhofer, McQuown, and Vasicek
LGD – Loss Given Default
OLS – Ordinary Least Square
PD – Probability of Default
PIH – Permanent Income Hypothesis
PP – Phillips-Perron
PSID – Panel Study of Income Dynamics
RR – Recovery Rates
SGS – Time Series Management System
SPC – Credit Protection Service
SVR – Support Vector Regression
VECM – Vector Error Correction Model

Chapter 1. Introduction

Brazil is the world's seventh wealthiest economy. It is also the largest country in Latin America and the Caribbean in terms of area and population. Brazil's economy outweighs that of all other South American countries, and Brazil is expanding its presence in world markets. Since 2003, Brazil has steadily improved its macroeconomic stability, building up foreign reserves, and reducing its debt profile by shifting its debt burden toward real denominated and domestically held instruments. After strong growth in 2007 and 2008, the onset of the global financial crisis hit Brazil in 2008. Brazil experienced two quarters of recession in 2009, as global demand for Brazilian commodity-based exports dwindled and external credit dried up. However, Brazil was one of the first emerging markets to begin a recovery. In 2010, consumer and investor confidence revived and GDP growth reached 7.5%, the highest growth rate in the past 25 years. Rising inflation led the authorities to take measures to cool the economy; these actions and the deteriorating international economic situation slowed growth to 2.7% in 2011, and 1.3% in 2012. Unemployment is at historic lows and Brazil's traditionally high level of income inequality has declined for each of the last 14 years. Brazil's historically high interest rates have made it an attractive destination for foreign investors. Large capital inflows over the past several years have contributed to the appreciation of the currency, damaging the competitiveness of Brazilian manufacturing and leading the government to intervene in foreign exchange markets and to raise taxes on certain foreign capital inflows.

Similarly, credit in Brazil has been growing very rapidly in recent years. Total credit to GDP has risen significantly in the last decade, by almost 25% points of GDP to about 45% of GDP. All credit categories have experienced strong growth rates but especially so consumer credit, which now represents 34% of total credit. A structural transformation has helped raise the supply and demand of credit. Capital inflows providing liquidity to banks, and the development of the domestic capital market, have fuelled the supply of credit. Economic stability, associated to a better business environment, strengthening labour markets and social mobility, have also raised the demand for credit by corporates and consumers.

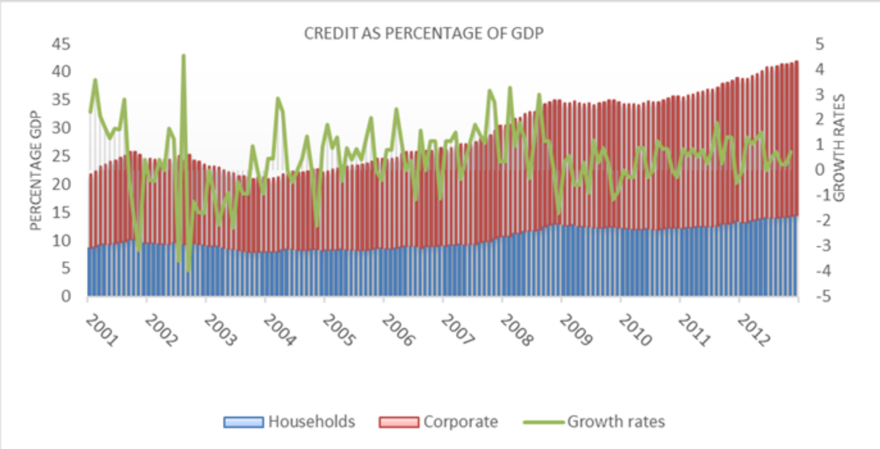


Figure 1 Brazil Bank Credit as Percentage of GDP

Credit risk assessment plays an important role in the credit risk decisions of financial institutions and it is crucial for financial regulatory issues, which have become more critical since the recent economic crisis, have prevented a large number of credit consumers from paying off their loans. For this reason, in 2006, the Basel Committee on Banking Supervision published the Basel II framework. The objective of the Basel II Accord is to better align the minimum capital required by regulators with risk. This inevitably requires a more complex regime, given that some of the greatest anomalies in the first Basel Accord stemmed from its simplicity. The Basel II Accord is based on three pillars: minimum capital requirements (pillar I); the supervisory process (pillar

II); and market discipline (pillar III). It implies not only bringing into line regulatory capital more adjusted to risk but also to promote a more sophisticated approach to risk management which means understanding risk and remaining alert to risk as a core issue. Brazil implemented the Basel II accord with the aim of applying it uniformly to all institutions in the Brazilian financial system.

1.1. Research Aims and Importance

The objective of this study is to investigate novel information in estimating Loss Given Default for personal loan portfolio in Brazil. This thesis explores three issues: first, in Chapter 3, it considers whether or not there is a short and/or a long-run relationship between the delinquency rates in Brazilian consumer personal credit and macroeconomic variables by applying VECM (Vector Error Correction Model), which adjusts to both short run changes in variables and deviations from equilibrium. Second, in Chapter 4, it investigates reasons for missing payments and establishes whether or not these reasons are related to a customer’s propensity to risk and the extent of their financial knowledge. This research was undertaken by surveying borrowers who were in arrears, combining answers from the questionnaire with borrower information as exploratory variables; and third, in Chapter 5, it explores the impact of collection actions in estimating LGD using Retail Loan Level Panel Data, which are activated by tracking customer payment following the collection process.

Assuming that there is a long-run relationship between Loss Given Default (LGD) and macroeconomic variables, these time series can be used in a Panel Data framework to improve the performance of the model. In the same way, if the reason for missing payment is shown to be significant in predicting recovery from default, this information can be captured by the collection process, which in turn can be of importance in establishing who is more likely to pay a loan off. Thus, this information can be used to improve the model accuracy.

Since this study investigates Loss Given Default for a Personal Loan portfolio from a Brazilian lender, it explores six Brazilian macroeconomic variables: Consumer Confidence Index, Aggregate Consumer Price Index, Personal Loan Outstanding Balance, Gross Domestic product and Unemployment Rate, which could explain Personal Loan Delinquency. In addition, two dependent variables were investigated namely Delinquency Rates from 15 days to 90 days and Delinquency Rates of more than 90 days. All of these time series are published by the Brazilian Central Bank.

1.2. Contributions

The existing literature in Loss Given Default (LGD) does not consider why people delay repayments or why they are in default. This research fills this gap by investigating the reasons for default and exploring ways of incorporating these reasons into recovery rates models, since the standard approach overlooks the real reasons for default and only uses proxies for them such as marital status and length of employment. This objective will be achieved by eliciting information through interview with customers who fail to repay their loans in order to discover the causes for this failure. In addition, the survey aims to consider the customer's personality traits and their financial knowledge in relation to the reasons for default. This information is then linked to the applicant's socio-economic and behavioural characteristics.

In addition, previous studies in this area do not associate the impact of the collection actions, over time, on the reduction of outstanding debt of delinquent borrowers. This thesis addresses this issue by developing models that use collection actions as input to predict their effects on recovery rates.

Moreover, one of the reasons why there is so little research that investigates recovery rates could be attributed to the fact that collections are usually managed apart from the portfolio system both in house and out of house. It is very challenging to grade data at account level which contains timelines of action linked to a borrower's details. In addition, even when a house procedure is run by the financial company there is no exchange of information between the collection process and portfolio management. Therefore, for improvements in the accuracy of recovery rate models at account level, it is crucial to acquire detailed collection data (De Almeida Filho, Mues, and Thomas 2010).

For this reason, the novelty of this study refer to building a retail level Panel Data model that integrates borrowers' application characteristics, changes in outstanding balances and collection actions for those who have missed payment at an early stage. This model estimates recovery rates based on the impact of collection actions that are taken at each stage of the process by tracking customer payments following collection actions.

1.3. Thesis Overview

The thesis proceeds as follows:

Chapter 2 presents a literature review on Loss Given Default (LGD) and Recovery Rates (RR) modelling approaches.

Chapter 3 estimates the aggregate delinquency in loan retail using Brazilian macroeconomic variables such as Growth Development Product (GDP), Consumer Confidence Index (CCI), and Interest Rates.

Chapter 4 predicts recovery rates taking into account borrowers' reasons for missing payment, risk-taking propensity and degree of financial knowledge.

Chapter 5 investigates the impact of collection actions on recovery rates. This is achieved by combining individual customer information and collection actions, over time, using retail loan level Panel Data.

Chapter 6 presents the concluding remarks, summarises the limitations of this study and proposes future research in the area.

Chapter 2. Prominence of Loss Given Default

In this chapter, the role played by Loss Given Default as an important parameter in credit loss is presented. In addition, a brief survey of previous studies in Loss Given Default is portrayed.

2.1. Bank Regulation

The Basel Committee on Bank Supervision (BCBS) was founded in 1974 as an international forum where members could cooperate on banking supervision matters. This committee was created in response to the crash of the stock market. The BCBS aims to enhance "financial stability by improving supervisory knowledge and the excellence of banking supervision worldwide." This improvement is activated through regulations known as accords Basel I, II and III.

Basel I is a set of international banking regulations published by the Basel Committee on Bank Supervision (BCBS), which determines the minimum capital requirements of financial institutions with the objective of minimizing credit risk. Banks are required to preserve a minimum capital based on a percent of risk-weighted assets. Basel I is the first of three sets of regulations recognised as Basel I, II and III and together as the Basel Accords. It was released in 1988 and it focused primarily on credit risk by producing a system for the classification of bank assets.

The Basel II Accord was announced after considerable losses in the international markets after 1992, which were accredited to weaknesses in credit risk management practices. The use of standardised measurements for credit risk, market risk and operational risk became mandatory.

In credit risk, capital requirement can be measured in function of the degree of sophistication. There are three possible approaches: the Standardised Approach, the Foundation Internal Rating-Based (IRB) Approach, and the Advanced IRB Approach. Banks are allowed to use their own internal measures for key drivers of credit risk as primary inputs to the capital calculation: PD (Probability of Default), EAD (Exposure at Default), and LGD (Loss Given Default), subject to meeting certain conditions and to explicit supervisory approval. The benefit in adopting the IRB is that financial institutions could be potentially recompensed with lower risk capital requirements.

Supervisory capital charges are calculated as Expected Loss (EL) of a portfolio which can be defined as the proportion of borrowers who will possibly default within a specific period (Probability of Default) multiplied by their outstanding balance (Exposure at Default), and multiplied by the percentage of their outstanding balance which will not be recovered (Loss Given Default). Intrinsically, the three factors mentioned above correspond to the risk parameters upon which the Basel II IRB approach is developed.

The Basel Committee specified a risk weight formula in which the risk parameters are transformed into risk weights and regulatory capital. Figure 2 shows an example of how losses of a portfolio can be recognised overtime.

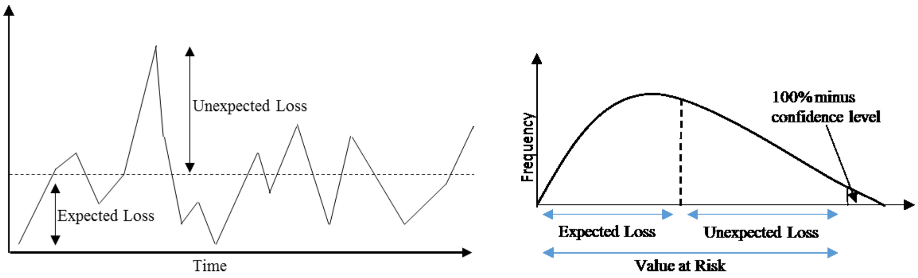


Figure 2: Value at Risk Components

Banks should hold capital based not only on the expected losses but they must also take into account the unexpected losses, which can be defined as the difference between Value at Risk and expected losses. The expected losses of a portfolio are built on the basis of three components: PD, EAD and LGD. This finally leads to the Basel IRB formula used to calculate the required regulatory capital to cover estimated unexpected loss:

Value at risk is the level of capital that is required to prevent the bank from going bankrupt in one year with the probability of no more than 100% minus de confidence level. Vasicek (1987) developed a model which was adopted by Basel IRB. This model formula is presented below.

$$VaR_i(\alpha) = \sum_{i=1}^{N} \overline{EAD_i} \bullet \overline{LGD_i} \bullet \Phi_N \left[\sqrt{\frac{1}{1-\rho}} \Phi_N^{-1}(PD_i) + \sqrt{\frac{\rho}{1-\rho}} \Phi_N^{-1}(\alpha) \right]$$

Where Φ is the cumulative standard normal distribution, α is the confidence interval and ρ is the asset correlation. The Vasicek (1987) work was based on a credit risk model developed by Merton (1974). Although Merton's model was exploring the value of a single firm, Vasicek was investigating probability of default on portfolio level (Thomas and Wang 2005).

According to Thomas and Wang (2005), Vasicek's model is applied to estimate PD in economic downturn. First, inverse cumulative standard normal distribution is applied to calculate probability of default. Similarly, a risk factor can be derived using the same process, which will help to predetermine supervisory confidence level. Thus, a downturn default threshold is produced by the correlation weighed sum of default threshold and the value of the single factor

With regard to the supervisory confidence level, which is fixed at 99.9%, it could be considered to be very conservative, however, Vasicek equation assumptions estimate reality. The high level of confidence is justified since its conservatism would compensate for the uncertainties of the Vasicek model (Committee 2005). Error estimation might occur in banks' internal PD, LGD and EAD models. Further, Vasicek's assumption is that a credit portfolio is infinitely fine-grained and a single risk factor is normally distributed, which, in reality very rarely occurs (Thomas and Wang 2005).

Therefore, PD and LGD modelling have become very interesting topics of exploration for researchers in credit risk management who wish to improve the accuracy of such models and support financial institutions in developing their own LGD models according to Advanced Internal Rating Based (AIRB) and consequently reducing their amount of capital that should be held.

IFRS 9 which is a new methodology for calculating expected credit losses, will be effective from 2018 and replaces the current IAS 39 approach. It essentially adjusts the classification and measurement of financial instruments. Under IAS 39, banks were allowed to recognise a credit loss on a financial asset, only when there was objective evidence that an impaired event had occurred. This method underestimates the required provisioning levels of banks, since it delays the recognition of credit losses. Arguably, this was one of the contributing factors of the credit crisis. The purpose of IFRS 9 is to increase financial instability by introducing a forward-looking expected loss impairment model, which allows banks to provision when a financial asset is recognised. IFRS 9 will have a significant impact on the risk modelling landscape of banks and those that are already Advanced Internal Rating Based (AIRB) compliant would make an easier transition. However, IFRS 9 requirements and definitions differ significantly and a considerable effort is required. Likewise, for banks that are not yet AIRB compliant, implementing IFRS 9 could be a springboard to AIRB compliance.

2.2. Loss Given Default Distribution

Several studies suggest that due to the data LGD is bimodally distributed and bound at 0 and 1. LGD tends to be distinguished by high concentration on the extremes representing either total recovery or total loss or both. Most of the empirical studies account a greater peak on zero and a smaller peak on one (Bastos 2010; Gupton 2005; Calabrese 2012; Caselli, Gatti, and Querci 2008; Chalupka and Kopecsni 2008). Caselli et al. (2008) observe the opposite: a large peak on one and a smaller pick on zero. Bellotti and Crook (2012) suggest similarly substantial peak on both zero and one for credit card portfolios. However, Gupton (2005) observes only a large peak on one for the corporative sector. Similar observations are obtained in LGD studies, whose applications are not focused on forecasting LGD (Araten, Jacobs, and Varshney 2004; Asarnow and Edwards 1995; Grunert and Weber 2005; Friedman and Sandow 2003; Renault and Scaillet 2004). Based on these studies, the connection between the relative size of the peaks on zero and one and the type of portfolio does not seem to be obvious. These differences may be triggered by factors such as internal bank policies or external economic conditions.

2.3. A Brief LGD Survey

One of the earliest studies in credit risk parameter modelling was Merton (1974). This model assumes that the probability of default of a company is determined by the value of its assets. Following Merton's work, Black and Cox (1976) considered more complex capital structure while Geske (1977) investigated the inclusion of interest on debt payment and Vasicek (1987) introduced the difference in short and long-term obligations which was the start of KMV models.

These models suggested that recovery from default is a function of the structural company's characteristics such as asset levels, business risk and financial risk. Consequently, recovery rate is the dependent variable, which is a function of the value by the defaulted firm's assets (Geske 1977; Black and Cox 1976; Merton 1974). According to Merton (1974) probability of default and recovery rates have a tendency to be in reverse association. Thus, when the value of a firm increases then its PD will decrease while RR at default will increase if all other parameters remain unchanged. On the other hand, if the debt of the company increases, its PD increases and the expected RR decreases. However, Franks and Torous (1994) implies that the lognormal distribution applied by the Merton model tends to inflate recovery rate in the event of default.

The following studies in this field aim to correct the main weakness of Morton's model which is the assumption that the event of default only take place if the company's assets are no longer sufficient to cover debts. As an alternative, it is expected that default could occur any time during the lifetime of the loan and it could materialise when a company's asset is lower than its net asset value (Hull and White 1995; Kim, Ramaswamy, and Sundaresan 1993; Longstaff and Schwartz 1995). These models assume that RR is a fixed ratio of the outstanding debt and it is not associated with PD.

Longstaff and Schwartz (1995) used history default data from Moody's corporate bond to more accurately estimate recovery rates. In addition, they found that there is a correlation between risk of default and interest rates, which has a substantial impact on credit spread indices.

Eom et al. (2004) investigate models of corporate bond pricing by testing the five structural models developed by Merton (1974), Geske (1977), Longstaff and Schwartz (1995), Leland and Toft (1996), and Collin-Dufresne and Goldstein (2001). Their findings suggest that accuracy is a problem and those models are likely to exaggerate the risk of credit of companies that have high weight volatility and also they under predict credit spread for good bonds. In contrast, the Leland and Toft model, however, over predicts.

There are two other problems that compromise the performance of these structural models: credit-rating changes before actually becoming defaulters and the assumption that the value of a firm is constant in time. With a view to improving these models, a new generation of models called reduced form models has emerged. Those of

Litterman and Iben (1991), Madan and Unal (1995), Jarrow and Turnbull (1995), Lando and Turnbull (1997), Lando (1998), Duffie (1998), and Singleton (1999).

These models do not estimate the company values since, for them, they are not associated with default. Moreover, PD and RR are built independently and they are based on defaulters' behaviour. The models are able to deal with unexpected defaulters, assuming that the probability of default over time is greater than zero. In addition, these models agree that recovery rate is random and subject to the event of default, and it could be correlated to macroeconomic factors (Singleton, Orthofer, and Lamuela-Raventós 1999; Jarrow and Turnbull 1995). These models, called reduced-form, vary in terms of parametrisation of the dependent variable recovery rate.

Credit Value at Risk models which are viewed as reduced form models, consider RR in a different way. For example, CreditRisk+ models, assume that RR is a constant parameter. On the other hand, CreditRisk models treat RR as a stochastic variable, independent from the probability of default (Gordy 2000).

2.4. Corporate Bonds LGD

In 2002, the company Moody's Analytics developed a methodology called LossCalc for predicting Loss Given Default (LGD). This methodology uses a dataset of 900 defaulter loan accounts, bonds and preferred stock from private and public firms in the United States, covering a period of 20 years. The model is able to estimate LGD for both recent default and default that would occur in one year's time. The developers claim that this model supports all the accuracy expected by BIS (Bank for International Settlements) though the Basel Capital Accord (Gupton et al. 2002).

Since then a number of researchers, who investigate risk for corporations, were dedicated to modelling PD but only few studies were conducted in modelling LGD. More recently, Jacobs et al. (2011) used Moody's recovery data and applied the beta-link generalised linear model to build a predictive econometric model for LGD. Their findings suggest that the robustness of their model relies on the fact that it can coherently interact LGD to an obligor's instrument level.

Likewise, accessing Moody's ultimate recovery database, Khieu et al. (2012), built a model to estimate recovery rates using loan information, borrowers' details, market conditions, economic factors, and recovery strategies. The results of this study advocate that loan information is more significant than borrowers' characteristics. Market and economic environments can be considered relevant when lenders offer a standard set of products, without differentiating customers. In addition, they conclude that a conventional procedure use of a 30-day post default trading price of the loan as a proxy for recovery rates is biased and ineffective.

Besides, the study conducted by Bastos (2014) produced a model using data from the Moody's company which contains US non-financial corporations that hold over 50 million dollars in default debt. A new data set was produced by bootstrapping observations and calculating the mean of the models' exploratory variables using a simple regression technique. After that, a decision tree induction algorism was applied. According to Bastos (2014), the results indicate that ensemble models seem to have more predictive power than a single model for bonds or loans across the entire range of recovery values.

Altman et al. (2014), also using Moody's ultimate recovery data set, proposed an approach to estimate recovery rates. This model employed tree regression based on a mixture of Gaussian distribution utilising conditioning information from borrowers' characteristics, loan information and credit strategy. They conclude that this model is shown to be more accurate than simple regression techniques. In addition, this methodology performs better than models that are developed using parametric regression.

Jokivuolle and Viren (2013) present an empirical model using a random-sampling method to simulate LGD annual average. This random time series sample contained information regarding credit risk of corporate bank loans from 1989 to 2008. In addition, the time series for the exogenous variables to capture downturn was applied with a view to stressing the models. This model is based on two equations: One for PD and the second for LGD. The empirical results suggest a positive relationship between PD and LGD and their cyclical movement response with a business cycle. Moreover, this study argues that the assumption of an exogenous LGD is not accurate.

Addressing LGD for corporate bonds, Yao et al. (2015) investigated Support Vector Regression (SVR) for forecasting the Loss Given Default of corporate bonds. A comparison of this methodology with thirteen other methodologies was considered. The results imply that the proposed algorithm performs better than those objects of the evaluation. From the results based on the bonds segmentation in function of their maturity, it could be concluded that least square SVR has more predictive powers. Moreover, it is important to emphasise that this approach, SVR, can be successfully applied for banks to predict Loss Given Default.

Employing the option pricing theory for modelling LGD for mortgage portfolios, Frontczak and Rostek (2015) argue that one of the benefits in applying their methodology is to reduce the complexity involved in modelling LGD for this type of portfolio. Another contribution is the inclusion of the type of property into the model which allows addressing the uncertainty related to the collateral. In addition, the introduction of liquidation costs can be seen as an advantage of this model (Frontczak and Rostek 2015).

Dermine and De Carvalho (2006) stated that more of research is conducted on LGD for corporate bonds than bank loans, which can be explained by the fact that bank loans are private instrument; consequently there are insufficient data available for the public. Their study aims to apply mortality analysis on defaulted bank loans of a European bank. They consider the value of the cash flow to be recovered once the event of default occurs as recovery rates. They argue that there are a number of exploratory variables associated with recovery rate such as the size of the loan, collateral, industry sector, and age of the firm. In addition, it can be supposed that the availability of data that contain recovery rates over time would allow the development of dynamic measures for provisioning loss (Dermine and De Carvalho 2006).

Insolvency by companies is another relevant subject for LGD modelling studies since it can be very difficult for them to recover from default. Dermine and De Carvalho (2008) investigated legal and internal collection action that was employed by these firms such as estimating recovery rates. They consider three types of legal action: foreclosure, provisional seizure, and injunction. In addition, internal action to promote cash collection was taken into account. The findings suggest that collection actions have predictive power and improve the accuracy of recovery rate models.

2.5. Personal Loans Portfolio LGD

Recovery rates for unsecured loans seem to be harder to estimate because of the absence of collateral. In addition, they rely on both the borrower's willingness to make their repayments and the lender's collection strategy (Matuszyk, Mues, and Thomas 2010) .

In light of the New Basel Accord, a study conducted by Zhang and Thomas (2012) aimed to compare recovery rate models using linear regression and survival analysis techniques. The data used allow the calculation of historic recovery rates. Their results show that linear regression outperforms models developed using survival analysis techniques. In addition, it was pointed out that the reason for the poor performance of survival analysis models could be the size of the data.

According to Matuszik (2010), the new Basel Accord published in 2007 significantly changed how financial institution developed credit risk models, particularly Loss Given Default (LGD) models. Thus, they suggest an approach which combines a decision tree and regression techniques to estimate recovery rates (RR). This study used data from a UK lender which contains information about defaulters who were taken into the collection process; accounts were classified into two categories: those which were recovered either partially or totally, and those which were not recovered. The exploratory variables used are loan amount, months in arrears, time at current address and whether or not the loan had joint application.

The results suggest that the in-house collection strategy is straightforward but the outside house collection process should be investigated. Moreover, this study confirms the necessity of the inclusion of a macroeconomic scenario into the model with a view to capturing customer behaviour in downturn conditions.

Hoechstoetter et al. (2012) investigate recovery rates for personal loan debt retail credit which was purchased by a German collection company. The model was built in two stages. In the first stage defaulters were classified into two categories, those who paid in full or did not pay any amount and secondly those accounts that showed a degree of payment. Those accounts which have recovery rate greater than zero and less than one will be used in the second stage with a view to estimating RR. Statistical and non-statistical techniques were applied and the findings suggest that the non-statistical approach performed better than the statistical one.

Thomas et al. (2012) investigated whether or not there were differences between in-house and collection agents with regard to their procedures. The data of defaulters from unsecured personal loan portfolio were used. The outcomes of the model imply that the two-stage model proposed can be applied to both in-house and to a third party in order to calculate the value of the debt for buying/selling purposes. Moreover, collection departments could use the model to define time to sale debt.

Using credit card data from a lender based in the UK, Bellotti and Crook (2012) built a number of Loss Given Default models using different statistical techniques such as Tobit regression, Ordinary Least Square (OLS) regression and Decision Tree. Customer details at the time of application, applicant personal lifetime affordability, account information, macroeconomic scenario at the time of the event of default, and lender credit strategy can be listed as the types of variables used into the model. Based on their findings, they concluded that default balance is associated with recovery rates and the use of macroeconomic variables require data across the business economic cycle. In addition, OLS is shown to be the best technique for forecasting LGD.

The internal ratings-based (IRB) approach requires that estimates should reflect economic downturn conditions to capture risk accurately. With a view to building an LGD model in accordance with Basel requirements, Calabrese (2012) proposed a regression methodology for modelling recovery rate as a mixed random variable expressing the extremes through a mixture of models developed by Bernoulli. The continuous elements of recovery rates (greater than 0 and less than 1) were assessed by beta regression. Although the model allows the analysis of the effect of the same covariates on the extreme values it is not able to replicate multimodality in the interval between zero and one.

According to Gürtler and Hibbeln (2013) Loss Given Default for bank loan portfolios is based on historic average if the information used relating to a particular debtor’s status is overlooked, which is very important from a regulation perspective.

Using data from German bank loans, Memmel et al. (2015) argue that industrial composition, maturity structure, regional factors, and exposure to the global economy scenario have a significant impact on bank credit portfolio losses.

According to Thomas et al. (2009) modelling LGD based on a collection process is particularly relevant for unsecured loans where repayment depends both on a customer’s capacity and inclination to make a payment and the banks’ strategy in approaching those who are in arrears. In addition, debtors are likely to oscillate between payment and non-payment over time. This work applies Markov Chain to modelling LGD using information from the collection process at portfolio level. The findings suggest that the model is not only able to predict LGD but also it underlines how LGD values depend on lender collection strategies. They conclude that models accessing collection information at account level to estimate debtor repayment behaviour should be investigated.

Chapter 3. Understanding Aggregate Default: An Empirical Investigation of Brazilian Loans Using Co-integration Vectors

3.1. Introduction

The 2008 global finance crisis is considered by many economists to have been the worst financial crisis since the great depression of 1929. This event has raised the need to improve delinquency rate forecasting.

In order to understand factors that could drive delinquency, firstly, it is important to identify its factors since delinquency occurs when the borrowers do not pay off their debt. This indebtedness is linked to consumption; levels of indebtedness of households, which increase when individuals spend more than their income. The literature in economics has a number of studies dedicated to consumer consumption behaviour and savings, however, there are few studies related to delinquency determinants and their effects on the business cycle.

The Permanent Income Hypothesis (PIH) is an economic theory developed by Friedman (1957) which tries to explain the consumption choices of agents over various periods of time. It supposes that an individual's consumption at decision in the present is determined not only by their actual income but also by their expected income in the future. The PIH assumption states that changes in permanent income drive the changes in a consumer's consumption behaviour. Therefore, if consumption is a function of the individual's expectation they could increase their indebtedness in the present, based on an optimistic view of the economy, culminating in delinquency if their supposition was not met. Consequently, consumer confidence could explain delinquency, which would be associated with the breakdown of consumer expectation.

Additionally, evaluating a consumer intertemporal restriction could originate delinquency. When individuals decide on consumption, they can also define the level of indebtedness, and consequently how much money they are prepared to spend on interest rates in the future. Therefore, variables that influence income, expenditure and interest rates for consumers could explain delinquency. In others words, variations in income or costs of living from fluctuations of business economic cycles could directly be linked to delinquency.

With a view to establishing the factors that could take households to delinquency, Stiglitz and Weiss (1981) introduced a credit-rationing model. This model assumes that lenders do not have all the necessary information regarding the ability of borrowers to repay, therefore the interest rate charged is contained within the risk they are taking. On the other hand, this additional value on top of the interest rates could influence lending to riskier customers, which would affect their delinquency rates.

This chapter explores the determinants of the aggregate delinquency of personal loan credit portfolios from Brazil. The selection of the variables was based on previous empirical studies and aimed to explain aggregate delinquency for 15-90 days of missing payment and for being more than 90 days overdue. It is of interest for this research to understand which macroeconomic variables could explain delinquency.

After the data collection of the time series believed to be associated with delinquency, two models were built, one for each stage of delinquency. The macroeconomic variables used were: unemployment rates, GDP, personal loan interest rate, aggregate price consumer index, personal loan outstanding balance, and Consumer Confidence Index.

Those variables are thought to be related to the hypothesis that a customer's ability to pay is determined by certain factors such as sensibility to credit cycles, future expectation not being met and changes in personal circumstances. A Vector Error Correction Model (VECM) was applied by means of estimating long-run and short-run relationships between those macroeconomic variables and personal loan delinquency.

3.2. Economic Factors

Macroeconomic conditions can provoke systematic changes that are central to credit risk assessment. Despite this fact, the literature focusing on the relationship between credit default and macroeconomic environment is rather sparse. Early studies explored the link between rating changes and macroeconomic conditions. Later studies that used cross-sectional or Panel Data methods include (Nickell, Perraudin, and Varotto 2000), (Bangia et al. 2002), (Zakrajsek, Carpenter, and Whitesell 2001) and (Kavvathas 2001). The first two of these later studies used GDP growth to classify the different phases of business cycles and compute separate default and rating transaction probabilities for each of these regimes. Kavvathas (Kavvathas 2001) applies a duration model for rating transactions and incorporates macroeconomic variables to capture systemic effects on transition probabilities. Studies that explore the time series approach include (Koopman and Lucas 2005) and (Koopman, Lucas, and Monteiro 2008). A multivariate unobserved components framework was employed to study cyclical co-movement between GDP and business failures. This research discovered evidence to support the premise that there is a relationship between credit risk and macroeconomic factors.

Other publications from this limited literature relate default correlations to macroeconomic conditions. Default correlation is a measure of interdependence among risks, and its own concept embodies the idea that common events (such as the business cycle) might lead default events to occur in clusters. Nagpal and Bahar (2001), for example, modelled default correlations and concluded that credit events are correlated and caused by common economic conditions. De Servigny and Renault (2002) investigated default correlation empirically and found higher coefficients for recessionary periods using data from U.S. companies. Cowan and Cowan (2004) utilised a large portfolio of residential subprime loans to demonstrate that default correlation is substantial in the data and that regulators and lenders would be well advised to develop more sophisticated credit measures. They also suggested that the impact of changes to the business cycle on portfolio losses should be incorporated into credit risk models. Trück and Rachev (2005) conducted an experiment using the Value at Risk approach based on a loan portfolio of a large European bank. The experiment revealed that losses were much higher in times of recession than during periods of economic growth.

More recently, after widespread concerns about the possible procyclical effects of the Basel II Accord on the economy, there has been a considerable flurry of activity around this theme. Koopman et al. (2008) discovered cyclical behaviour in default rates using a time series approach based on unobserved components and highlighted the main effects of this behaviour in a credit risk experiment, addressing the issue of procyclicality in ratings and capital buffer formation. Repullo and Suarez (2008) showed that banks have an incentive to maintain capital buffers, but that these buffers maintained in expansions are typically insufficient to prevent a contraction in the supply of credit in recessions. Repullo et al. (2010) compare alternative methods to mitigate the possible procyclical effects of the Basel II Accord. As a consequence of concerns about this issue, the Committee on Banking Supervision has begun to discuss the idea of capital buffers above the minimum regulatory capital of the banking sector during periods of large economic growth. Crook and Banasik (2012) argued that an increase in consumer delinquency would cause a decrease in banking sector profits and might increase the need to raise interest rate margins to compensate for losses from higher risk defaulters and, alternatively, institutions might have to increase their capital adequacy ratios. With a view to explaining aggregate delinquency, this study started from the premise that essentially there are two hypotheses that may explain why a borrower defaults, is unable to pay or has a default strategy. The findings suggest that macroeconomic conditions affect default rates.

3.3. Explaining Reason for Delinquency

Researchers continue to investigate the reasons why borrowers will possibly default on their debt. These studies follow two lines of explanation. The first is the hypothesis that customers are unsuccessful in meeting their obligations when unexpected circumstances occur such as job loss, marital breakdown or an increase in interest rates. A further hypothesis is that of strategic default, which takes place when a loan is collateralised by real assets because of a change in the economic environment, where the value of these assets becomes lower than the outstanding balance of the loan. This could compel the borrowers to cease making their remaining payments (Kau and Keenan 1995).

While investigating the impact of macroeconomic variables on aggregate delinquency, Ji (2004) accessed information from Panel Study of Income Dynamics (PSID), a longitudinal survey of randomly sampled American citizens and the houses in which they live. This survey has been conducted by Michigan University since 1968, and concentrates on dynamic aspects of household economic behaviour as well as demographic behaviour. Ji's study applied a probit technique to investigate whether or not unemployment rates influence consumer default. The findings suggest that default rates could be reduced by 30% when there is a stability in customer income over time, meaning that if there were an increase in unemployment rates, it would negatively affect default rates.

Furthermore, Gross and Souleles (2002) used Panel Data sets of credit card accounts to estimate duration models. The findings suggest that default increases with unemployment and low house prices, whereas large credit lines are more unlikely to default.

A further study was conducted by Dey et al. (2008) using data from a household debt survey in Canada. The research aimed to explain the relationship between customer indebtedness levels and delinquency. This was accomplished by applying simulation techniques on the aggregate Debt Service Ratio (DSR), which is the measure of a household's financial obligation. This process was employed in evaluating the impact of DSR on the probability of default. The results suggest that if the DSR reaches a threshold above 40%, it could indicate a prominent increase in the probability of default. Moreover, they concluded that a decrease in a household's income will significantly influence the Debt Service Ratio (DSR).

Ali & Daly (2010) applied the Logit technique to analyse the impact of macroeconomic variables on aggregate default levels, where they compared corporative and consumer credit data from the US and Australia. The results suggest that the Gross Domestic Product (GDP), short-term interest rates and total debt could explain the default rates of a country, which is consistent with the ability-to-pay hypothesis.

Lambrecht et al. (1997) applied survival analysis to differentiate between a defaulting borrower's ability to pay and strategic default. According to their findings there is more evidence of ability to pay than strategic default. While Deng et al. (1996) suggests that default can be triggered by divorce and unexpected unemployment.

With regard to macroeconomic determinants and other microeconomic delinquency, Crook and Banasik (2012) explored the co-integrated relationship between consumer delinquency and economic variables, using the Vector Error Correction Model (VECM) on a US dataset. Their findings show that the higher the interest rates the greater the delinquency.

3.4. Data Description

The times series, which relate to delinquency, personal loan interest rate, personal loan outstanding balance, and Gross Domestic Product (GDP) were gathered from the time Series Management System (SGS) , which is updated monthly and is based on the national finance system of the Brazil Central Bank. The time series data regarding unemployment rates was collected from the Brazilian Institute of Geography and Statistics (IBGE).

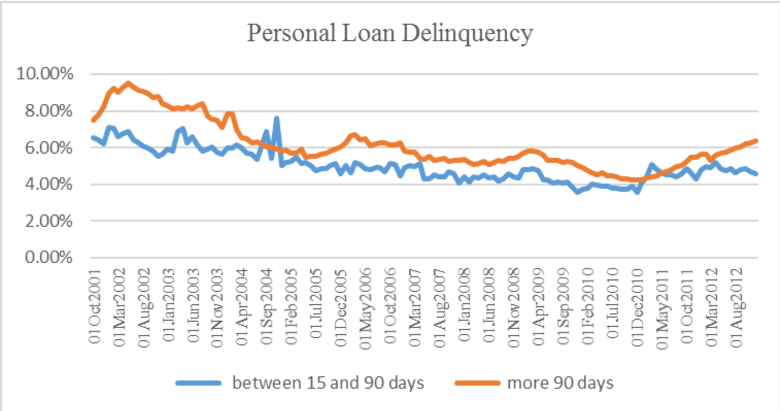


Figure 3 Personal Loan Delinquency Rate over Time

Figure 3 shows personal loan monthly delinquency rates from the Brazilian financial system between October 2001 and December 2012. During this period the performance of the personal loan portfolio demonstrated different behaviour. After reaching a peak of 9% in around April 2002, delinquency rates greater than 90 days presented a decrease until February 2005 after which they increased steadily until May 2006. Thereafter they fluctuated until June 2008 when the delinquency rates for more than 90 days reached 6%. There then followed another period of increase to December 2010. From that time they showed a constant increase and by December 2012 a rate of approximately 6.5% had been reached.

Delinquency rates between 15 days and 90 days behaved similarly to delinquency rates of more than 90 days except for the period between May 2002 and May 2003 when they had an opposite tendency and also from April 2004 until February 2005 when they peaked. Furthermore, after March 2012, while delinquency rates of more than 90 days increased, delinquency rates between 15 days and 90 days fell.

Crook and Banasik (2012) separated delinquency by portfolio type: credit card, real estate and personal loan. This study, however, investigates aggregate delinquency between 15 and 90 days and more than 90 days, since the objective is to understand whether or not there are different explanations for various lengths of missing payments. The explanatory variables that were tested to explain delinquency are: unemployment rate (unempr), personal loan interest rate (plir), log of personal loan deflated outstanding balance (lplob), log of deflated Gross Domestic Product (lgdp), Consumer Confidence Index (CCI) and Aggregate Consumer Price Index Adjusted (AIPCA). The variables lplob and lgdp were deflated using the monthly variation of market price index. Figure 4 shows the list of the variables that are used into the model.

Personal Loans Delinquency types:	**Name**
Delinquency rates between 15 days and 90 days	AR15D
Delinquency rates of more than 90 days	AR90D
Exploratory Variables	
Consumer Confidence Index	CCI
Aggregate Consumer Price Index	AIPCA
Personal Loan Interest Rates	PLIR
Personal Loan Outstanding Balance	LPLOB
Gross Domestic Product	LGDP
Unemployment Rate	UNEMPR

Figure 4 Variables Used into the Model

3.5. Model

Since this study is interested in exploring whether or not there is a short or long-term relationship between Brazilian macro-economic variables and personal loan aggregate delinquency, an econometric model was developed using VECM (Vector Error Correction Model).

When two or more series are individually integrated but some linear combinations of them have a lower order of integration, then the series are said to be co-integrated. In such cases a long-run relationship between these variables exists. The existence of a long-run relationship also has its implications for the short-run equilibrium relationship. This mechanism is modelled by an error-correction mechanism, in which the equilibrium error also drives the short-run dynamics of the series.

Assuming that there is a long-run linear relationship between personal loan aggregate delinquency and its determinants, these macroeconomic variables can be utilised for forecasting Loss Given Default over time.

This experiment was achieved by using time series data published by the Brazilian Central Bank, from October 2001 to December 2012. Each set of time series variables was tested for unit root using the Phillips-Perron and Dickey-Fuller tests. For those variables that have unit root, Johansen's technique will be applied. If a set of variables are found to have one or more co-integrating vectors then a suitable estimation technique is a VECM (Vector Error Correction Model), which adjusts to both short-run and long-run changes in the dependent variable. An important econometric task is in determining the most appropriate form of trend in the data. For example, in ARMA modelling the data must be transformed into stationary form prior to analysis. If the data are trending, then some form of trend removal is required. Two common trend removal or de-trending procedures are first differencing and time-trend regression. First differencing is appropriate for I (1) time series and time-trend regression for trend stationary I (0) time series. Unit root tests can be used to determine if trending data should be first differenced or regressed on deterministic functions of time to render the data stationary on equilibrium (Patterson 2000).

Moreover, economic and finance theory often suggests the existence of long-run equilibrium relationships among non-stationary time series variables. If these variables are I (1), then co-integration techniques can be used to model these long-run relations. Hence, pre-testing for unit roots was applied as a first step in the co-integration modelling process.

The theory behind ARMA estimation is based on a stationary time series. A series is said to be (weak or covariant) stationary if the mean and auto-covariance of the series do not depend on time. Any series that is not stationary is said to be non-stationary. A common example of a non-stationary series is the random walk:

$$Y_t = Y_{t-1} + \varepsilon_t$$

where ε is a stationary random disturbance term. The series Y has a constant forecast value, conditional on t, and the variance increases over time. The random walk is a different stationary series since the first difference of Y is stationary:

$$Y_t - Y_{t-1} = (1 - L)Y_t = \varepsilon_t$$

A difference stationary series is said to be integrated and is denoted as $I(d)$ where d is the order of integration. The order of integration is the number of unit roots contained in the series, or the number of differencing operations it takes to make the series stationary. For the random walk above, there is one unit root, so it is an I (1) series. Similarly, a stationary series is I (0).

Standard inference procedures do not apply to regressions that contain an integrated dependent variable or integrated regressors. Therefore, it is important to check whether a series is stationary or not before using it in a regression. The formal method to test the stationarity of a series is the unit root test.

The Dickey-Fuller test is a methodology to determine whether the above process has a unit root. Initially, the first difference of a time series variable is calculated by

$$Y_t - Y_{t-1} = \rho Y_{t-1} + \varepsilon_t - Y_{t-1}$$

If we use the delta operator, defined by $\Delta Y_t = Y_t - Y_{t-1}$ and set $\beta = \rho - 1$, then the equation becomes the linear regression equation:

$$\Delta Y_t = \beta Y_{t-1} + \varepsilon_t$$

Where $\beta \leq 0$ and so the test for ρ is transformed into test that the slope $\beta = 0$ parameter. Therefore, there are two hypotheses to be tested: $H_0 : \beta = 0 \quad (\rho = 0)$ a $H_1 : \beta < 0 \quad (\rho = 1)$ that is to say that if the ADF test statistic is positive, one can automatically decide not to reject the null hypothesis of a unit root.

Phillips and Perron (1988) developed their Phillips-Perron (PP) unit root test. Phillips and Perron's test statistics can be viewed as Dickey–Fuller (1979) statistics that have been made more robust to serial correlation by using the Newey–West (1987) heteroscedasticity auto correlation consistent covariance matrix estimator.

Although the PP unit root test is similar to the ADF test, the primary difference is in how the tests each manage serial correlation. Where the PP test ignores any serial correlation, the ADF uses a parametric autoregression to approximate the structure of errors. Both tests typically end with the same conclusions, despite their differences.

$$Y_t = \pi Y_{t-1} + \mu_t$$

In (1) μ_t is I (0) and may be heteroskedastic. The PP test corrects for any serial correlation and heteroscedasticity in the μ_t error non-parametrically by, modifying the Dickey-Fuller test statistics.

Under the null hypothesis, $\rho = 0$ the Phillips-Perron Z_t and Z_π statistics have the same asymptotic distributions as the ADF t-statistic and normalized bias statistics. One advantage of the Phillips-Perron test over the ADF tests is that the Phillips-Perron test is robust to general forms of heteroscedasticity in the error μ_t term. Another advantage of Phillips-Perron is that there is no need to specify a lag length for the regression test.

3.6. Results

Initially, Phillip-Perron test was conducted to check if the variables were stationary. It was assumed that a time trend for levels and without trend for first difference. This test is a precondition for Johansen Test of Co-integration in which all variables must be non-stationary at level but when they are converted into first difference, they will become stationary, meaning that all the variables are integrated of same order.

Table 1 contains the results for the Phillips-Perron unit root test, which shows that most of the variables were integrated at order 1 and consequently their first differences were stationary. The only exceptions were delinquency between 15 and 90 days (LAR15D) and Consumer Confidence Index (LCCI). Therefore, these two variables could not be used in the Johansen cointegration test.

Phillips-Perron unit root test					
		Levels	Adjusted t-statistic (with Trend)	Differences	Adjusted t-statistic (without Trend)
Personal Loans Delinquency types:					
Delinquency between 15 days and 90 days	LAR15D	-4.021	DLAR15D	-26.932	
Delinquency of more than 90 days	LAR90D	-0.686	DLAR90D	-10.279**	
Exploratory Variables					
Consumer Confidence Index	LCCI	-3.477	DLCCI	-15.297	
Aggregate Consumer Price Index	LAIPCA	-4.027	DLIPCA	-4.681**	
Personal Loan Interest Rates	LPLIR	-2.507	DLPLIR	-9.429**	
Personal Loan Outstanding Balance	LPLOB	-2.18	DLPLOB	-5.421**	
Gross Domestic Product	LGDP	-2.448	DLGDP	-4.674**	
Unemployment rate	LUNEMPR	-2.769	DLUNEMPR	-12.842**	

*Note: Significant at 5% = * significant at 1% = ***

Table 1 Phillips-Perron Unit Root Test

Table 2 demonstrates the Johansen cointegration test, which was only applied to the dependent variable LAR90D because the other dependent variable, LAR15D, was stationary at level. Trace statistic and Maximum Eigenvalue statistics reject the hypothesis of being stationary at most zero cointegration relationship but not that there are at least two cointegration equations at 5% level.

H_0 :	Trace statistic	5% cv	Max-Eigenvalue Statistic	5% cv
Personal Loans				
Delinquency rates of more 90 days (AR90D)				
$r = 0$	129.67**	95.75	46.83**	40.08
$r \leq 1$	82.84**	69.82	36.86**	33.88
$r \leq 2$	45.98*	47.86	19.42	27.58
$r \leq 3$	26.57	29.80	18.16	21.13
$r \leq 4$	8.40	15.50	8.19	14.26
Lags in ECM = 3				

*= significant 5% ; **= significant at 1%

Table 2 Johansen Cointegration Test

Table 3 shows that there is a long-run causality from aggregate Consumer Price Index, personal loan interest rates, personal loan outstanding balance, GDP and unemployment rate to delinquency rate of more than 90 days for personal loan portfolio.

Dependent variable (delinquency of more than 90 days)		Personal Loans (LAR90D)
Exploratory Variables		
Aggregate Consumer Price Index	LAIPCA	0.06613
		(0.37699)**
Personal Loan Interest Rates	LPLIR	0.03019
		(0.24502)**
Personal Loan Outstanding Balance	LPLOB	0.11312
		(0.24502)**
Gross Domestic Product	LGDP	1.1975
		(0.63618)**
Unemployment rate	LUNEMPR	0.93351
		(0.31781)**
Constant		-29.76601**

*= significant 5% ; **= significant at 1%

Table 3 Cointegration Vectors (normalised)

Regarding short-run relationships, Table 4 indicates that lag 3 of the dependent variable LDAR90D, lag 3 of DLUNEMPR and lag 1 of DLPLOB affect the delinquency rate of more than 90 days (LAR90D). There is no short-run causality running from the remaining variables to the dependent variable.

Dependent variable (delinquency of more than 90 days)	Personal Loans (LAR90D)	
	Coefficient	t_statistic
Exploratory Variables		
Δ Dependent variable		
DAR90D(-1)	0.077235	0.912915
DAR90D(-2)	-0.073302	-0.851843
DAR90D(-3)	0.330136	3.818908***
Δ Aggregate Consumer Price Index		
DLAIPCA(-1)	-1.264075	-1.214769
DLAIPCA(-2)	-1.221922	-0.941165
DLAIPCA(-3)	0.424166	0.381765
Δ Gross Domestic Product		
DLGDP(-1)	-0.740654	-0.941612
DLGDP(-2)	0.491847	0.546455
DLGDP(-3)	-1.143744	-1.527912
Δ Personal Loan Interest Rates		
DLPIR(-1)	0.029493	0.326509
DLPIR(-2)	0.048457	0.542876
DLPIR(-3)	0.007094	0.080579
Δ Personal Loan Outstanding Balance		
DLPLOB(-1)	-0.475128	-1.957129*
DLPLOB(-1)	-0.367698	-1.429734
DLPLOB(-1)	0.149522	0.632563
Δ Unemployment rate		
DLUNEMPR(-1)	0.109932	1.173503
DLUNEMPR(-2)	0.022808	0.249705
DLUNEMPR(-3)	0.209113	2.326104**
Constant	0.023505	2.307584**
Error Correction		
AR90D	-0.83968	-2.847746***
R^2	.0255976	
Adjusted R^2	.0128620	

Table 4 Short-run Dynamic Equations

The adjusted R^2 reveals a relatively low model fit which is corroborated by the plot shown in Figure 5. This means that the exploratory variables used in this study do not greatly help in understanding aggregate delinquency of more than 90 days in personal loan portfolios.

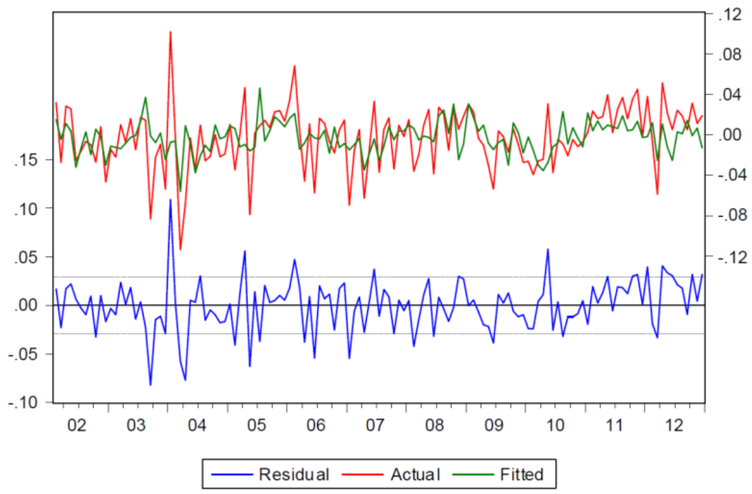


Figure 5 Model Fit

3.7. Conclusions

This study aimed to contribute in explaining aggregate delinquency rates between 15 and 90 days applying Vector Error Correction Model (VECM). Nevertheless, the results indicated that this dependent variable is stationary at level, which is one of the restrictions of the VECM methodology, namely variables should be non-stationary at level and stationary at first difference.

Another objective was to understand the aggregate delinquency rate of more than 90 days, using the aforementioned method. This variable complies with the method requirements.

In the long term, all the exploratory variables considered have influence on the delinquency rate of more than 90 days. On the other hand, in the short term only three variables have association with this delinquency rate: lag 3 of the dependent variable

LDAR90D, lag 3 of DLUNEMPR and lag 1 of DLPLOB affect the delinquency rate of more than 90 days (LAR90D). The unemployment rate lag 3 has association with the delinquency rate of more than 90 days, which is to be expected.

Although the model suggested a long-run relationship for LAR90D, the Adjusted R^2 indicates that the explanatory power of the independent variables is low.

Further work could be done to investigating new times series variables that have been published by the Brazilian Central Bank since 2013 which replaced those used into the models since this macroeconomic series were discontinued in December 2012.

Chapter 4. Estimating Loss Given Default taking into Account the Reasons for customers' non-payment

4.1. Introduction

Nowadays, lenders are more interested in meeting their business objectives such as profitability and market share than just forecasting default risk. They aim to optimize all their lending decisions by, for instance, choosing the credit limit, the interest rate and other product features that are offered to the customer so as to maximize their profitability. Moreover, the way the lender manages the relationship with the borrowers will affect their profitability, while the consumers are more conscious of their choices and are looking for more attractive products.

Customers are unsuccessful in repaying their loans for a number of reasons, ranging from simple forgetfulness to deliberate attempts. There are three main categories of customer in terms of missing payment behaviour: first, there are those customers who want to pay but fall into arrears owing to mismanagement or forgetfulness. Second, those who may also have good intentions but who are unable to meet their financial obligations because of unforeseen circumstances such as unemployment and marriage breakdown. Finally, the third category includes those who have the ability to pay but have no intention of doing so (Finlay 2008).

Thus, the purpose of this chapter is to fill this gap in the existing literature by investigating the reasons for borrowers' missing payment, customer propensity for taking risks and their knowledge regarding finance, and then using this new information to estimate recovery rates and consequently LGD.

This is achieved by modelling LGD using two alternative approaches: Ordinary Least Square (OLS) and Zero One Inflated Beta regression. The former is chosen because it is considered to be the most accurate methodology to estimate LGD (Bellotti and Crook 2012). The latter is considered since it is known for its effective fit to the LGD U-shaped distribution (Ospina and Ferrari 2012).

In this section, a brief overview of the previous studies which support the research is presented.

4.1.1. Recovery Rates

Recovery rate (RR) can be defined as the amount of an outstanding balance in default that can be recovered. Usually, it is expressed as a percentage of the amount's value (Bennett, Catarineu, and Moral 2005). It is associated with Loss Given Default (LGD), which can be expressed as LGD = 1- RR. Recovery Rate is measured in the interval between 0 and 1, where 1 means that the debt was fully recovery in total and 0 indicates no recovery. Values greater than zero and less than one, show that only part of the debt was recovered. Therefore, RR follows a bimodal distribution with frequencies that gradually decrease and then progressively increase. Usually, the frequencies are higher at the extremes of the distribution and lower in the centre. Recurrent and curved measurements are usually distributed in U-shapes (Bucher 2012).

4.1.1.1. Default

Loss Given Default is the portion the percentage of exposure the bank could lose if an obligor defaults. Consequently, LGD is intrinsically linked to the definition of default. The Basel definition of default is based on both a subjective and an objective condition.

A default is defined as the occurrence when the borrower is past due more than 90 days on a debt to the lender or when the lender considers that the borrower is unlikely to pay its balance (II 2004).

4.1.1.2. Previous Studies

Research has been conducted in modelling LGD throughout RR. However, these studies have been focused on assessing the probability of recovery by using certain personal characteristics available from the application form. Regarding profitability, one aspect that has a huge impact on it is the loss from debts that are written-off due the customers failing to repay their loans.

A number of previous studies estimate RR in different contexts. Altman et al. (2005), for example, applied a regression technique to investigate the relationship between recovery rates and economic scenarios, and between a loan's characteristics and default. Motivated by the Basel Accord, Lucas (2006) was concerned about the amount of outstanding debt that would be recovered on secured loans assuming that there is a gap between the collection state and the repossession of collateral assets.

Matuszik et al., (2010), established that the recovery rate is determined by both a customer's ability to pay back and the lender's collection strategy. This study applied linear programming to support lenders in deciding the best way to collect the debt, but it was not focused on the customer's side.

According to Bellotti and Crook (2012), a considerable increase in delinquency may cause lenders with low capital adequacy ratios to become insolvent. For this reason, decisions need to be taken about how to manage the delinquency so that the likelihood of the account recovering is maximized and potential future losses due to write-off are minimized.

4.1.2. Economic Psychology

Economic Psychology is an emergent discipline, informed by two disciplines (economics and psychology), which attempts to gain a greater understanding of how people behave in their economic lives within society. In order to comprehend economic processes, psychological factors need to be taken into account in terms of variables that express human motives, attitudes and aspirations (Katona 1975).

Despite Economic Psychology's original development being applied to marketing as a tool for understanding consumer behaviour, it has since extended its scope to focus on human behaviour associated with macro-economic matters such as saving, credit, income and inflation. The principal objective for psychology in economic research is to discover and analyse the forces underlying economic actions, decisions, and choices (Katona 1975).

According to Lea et al. (1995), Economic Psychology draws a distinction between credit, debt and indebtedness: where credit refers to an agreement between a lender and a consumer to pay back the outstanding amount. Concerning debt and indebtedness, the former occurs when a borrower delays making a payment, while the latter refers to the inability to pay off a debt.

For example, consumers who have spent money on foreign holidays, electronic equipment and vehicles express their lifestyle expectations, perceptions of luxury and needs. Furthermore, patterns relative to a person's perceived position in society, how they manage their debt and family finances are all in competition. In addition, the availability of credit, and an optimistic outlook have an impact on a person's choices.

With regard to debt, such problems lead to marital stress, depression and feelings of inadequacy. Reasons behind personal debt are inevitably of interdisciplinary interest: economics addresses the effects of income and life cycle models, sociology regards debtors as social groups and psychology seeks to understand the importance of people's behaviour and values within society.

Although a wide range of factors which impact on personal borrowing have been suggested, a conceptual model has yet to be proposed which incorporates these factors and research has tended to investigate relatively few of them.

A number of studies of reasons associated with personal debt have emerged. A number have investigated debt as the result of an inability to meet their financial commitments (Berthoud and Kempson 1992); (Baldwin and Ford 1988); (O'sullivan et al. 1989) (O'sullivan et al. 1989); (Livingstone and Lunt 1992). Furthermore, these studies have suggested that debt is more common among those with lower incomes and higher expenditure.

In contrast, other research has also indicated that credit is often seen as a sign of wealth and success, and that those families with higher incomes are inclined to higher instalment debt, (Katona 1975); (Cameron and Golby 1991).

A clear outcome of the studies reviewed here is a strong relationship between debt and social and psychological factors. Thus, falling into debt is not purely dependent on aspects of economic management.

The list of social and psychological factors that have been found or claimed to be correlative to debt is quite extensive. This study will consider a number of them namely: Locus of Control, Economic Risk Taking, The Big Five factors and Financial Literacy. Locus of Control was selected because it is an essential element in the understanding of the impact of the personality on a person's approach to credit (Livingstone and Lunt 1992; Tokunaga 1993; Davies and Lea 1995; Lea, Webley, and Walker 1995; Norvilitis, Szablicki, and Wilson 2003; Norvilitis et al. 2006b, 2006a; Perry 2008; Vio 2008). Risk-taking is included since it is significant in personal financial decision making (Blais and Weber 2009). Finally, research findings suggest that financial knowledge or lack of it is associated with an individual's personal credit profile (Perry 2008).

4.1.3. Risk-Taking

From the point of view of psychology, risk can be defined as a subjective construct of the interpretation of an event (Rottenstreich and Tversky 1997). Risk is also seen in terms of context (Diacon 2004). Research indicates that objective assessment has little impact on how people make investment decisions (Capon, Fitzsimons, and Prince 1996). Other aspects are thus taken into account when making decisions under uncertainty, and people are more influenced by perceived than objective risk (Diacon and Ennew 2001). Risk perception is an indispensable component of financial decision making and other risk-taking behaviour.

In addition, it has been noted that those who were reluctant to take a risk in the past are likely to continue to make cautious decisions, whereas those who have accepted risk in the past will probably continue to do so (Wallach and Kogan 1965). According to Einhorn and Hogarth (1981) people tend to attribute success to themselves and blame others or circumstances for their failure.

Risk-taking is a significant aspect of financial decision making and involves a flux of factors such as risk perception and risk attitude. What is more, it can be modified by socio-demographics and personality. People classified as being sensation seeking, extravert, and open to experience will probably take greater financial risks than those who are conscientious, anxious, or neurotic (Ding, Chang, and Liu 2009).

4.1.4. Financial Knowledge

Finance knowledge can indicate to what extent an individual can comprehend and apply personal finance-related information. Financial literacy can be employed in financial activities to increase expected lifetime utility from consumption. A financially literate individual might not display forecasted behaviour or increases in financial well-being because of other influences.

Lusardi ((2008) argued that many individuals in the United States and outside of the country do not comprehend basic financial concepts, for example, interest compounding rate, inflation, how assets are priced and variation in risk. The manner in which households make financial decisions could be associated with their lack of financial knowledge.

Perry (2008) considered the impact that financial knowledge had on personal credit risk. This study takes into account the FICO ratings in the USA as independent variables and a number of dependent variables, which help to explain customer behaviour. Two such variables were level of financial knowledge and locus of control.

The findings indicate a person who has high levels of financial knowledge and an internal locus of control will have a high credit score. In addition, the level to which financial knowledge influences a credit score depends on a person's personality.

Regarding indebtedness, Lusardi and Tufano (2009) conducted a study aimed at measuring debt literacy by applying a questionnaire that tested individuals' financial knowledge. They concluded that people with a lower level of financial literacy are more likely to pay higher interest rates and fees. In addition, as regards credit card expenses, they established that one third of the fees and charges paid by the customers could be attributed to lack of financial knowledge.

Moreover, Lusardi and Mitchell (2014) argued that in contrast to the microeconomic models, which usually assume that individuals are able to understand difficult monetary calculations and have skills to deal with financial markets, few people seem to have such financial knowledge.

Financial literacy has been regularly used as an alternative expression for financial knowledge, but in fact, their concepts are different. Huston (2010) states that financial literacy can be assumed as dividing into two dimensions: understanding and use. The first is related to personal finance knowledge and the second to the application of this knowledge.

4.2. Data Description and Experiment Design

The experiment draws on a data set from a store card issued by a Brazilian store retailing electrical appliances, electronic equipment and furniture. The store was founded in the 90s and currently has 68 branches. The data set is composed of 90,061 customers (see Table 5) and covers the period from April 2011 to July 2013.

	Customers	Loans	Instalments/Payments
Population	90,061	146,255	1,341,343
Sample	19,569	21,731	132,020

Table 5 Loan Data

The data contain monthly information relating to applications, customers and loans granted, and repayments. Preliminarily, the variables selected are listed on Table A1 in the Appendix. However, there are a large number of additional, potentially relevant variables that could be considered in the future.

A credit limit is given when the card is issued and the customer is free to make as many purchases as they wish (up to the credit limit). These purchases are repaid in fixed monthly instalments and the minimum repayment is the sum of the monthly instalments of each purchase. The customer is considered delinquent after 12 days of missing payment when they are approached by the collection team. However, after 45 days, if they have still not paid, their relevant details are sent to collection companies and are also recorded on the negative bureau data (SPC).

The data provider collection process is based on the length of time during which payments are missed. The first stage of the process involves reminding borrowers of their needs to make a payment; this contact is made by telephone after 5 days. The second communication is made by letter and this step is taken after a further period of 12 days without payment having been made. If payment is still not made after a further 25 days, a move is made to the third stage, when the borrowers' details are given to the negative bureau and a letter is sent to inform the borrowers of the action taken. The fourth stage in this process consists of generating data relating to those who have delayed in making payment for a total period of 35 days. This data is then passed on to an external collection company, who assume sole responsibility for collecting the payment. The final stage consists of writing off the contract with the assumption that these loans are lost.

This study is interested in investigating who becomes delinquent, why they do so and the causes of their subsequent move into default. The initial procedure consists of surveying customers that miss a payment. Data which contains the details of customers who became delinquent were sent to the survey company with a view to conducting a telephone interview. The customers were asked to complete a questionnaire that investigated the explanations given by respondents for their delinquency. In addition, it considered risk-taking attitudes and levels of financial knowledge which may contribute to understanding customer behaviour.

4.2.1. Observation Outcome window

Despite the availability data, the whole period was not used because it was necessary to take into account three factors which are essential for maximising the potential of the data: maturity, which means that the bad curve shows monotonous behaviour; censoring, which is related to the exclusion of the cases which go bad outside the window; decay, which means that all changes in the business should be considered, since it could bring bias to the model. According to (Siddiqi 2006), application scorecards tend to be built using an outcome window between 18 and 24 months for credit cards. This study has considered the outcome window of a 18-month period which is appropriated for a retail loan.

Figure 6 below shows the default rates in the data set by month. The percentage was calculated as a function of the number of loans that went to 45-day delinquency in each month and the number of loans that were active at that time.

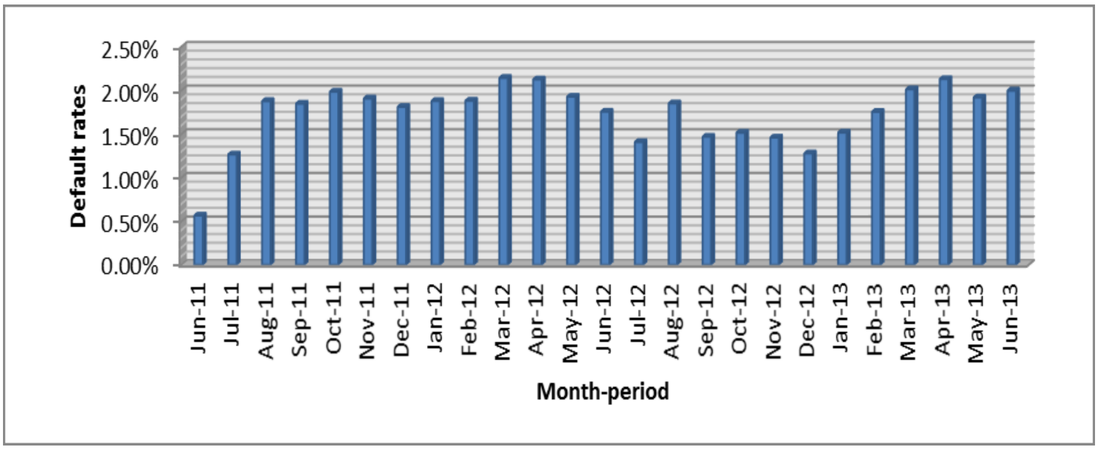


Figure 6 Cohort Analysis for Default Rates

4.2.2. Good/Bad Definition

The sample for this analysis refers to loans that were taken out from 1st September 2011 to 31st March 2012. Furthermore, loans that comprised of fewer than 12 instalments were excluded from the sample. In addition, instalments with a due date before 1st April 2012 and after 31st March 2013 were deleted, since the study aimed to observe a 12-month period. The sample was divided into two equal 6-month sub-periods designated ‘previous’ and ‘next’. The delinquency was calculated for each instalment of each loan. I have, however, only considered the worst delinquency for each sub-period. A transition matrix was then built based on the worst delinquency.

The matrix (see Table 6) shows that of the loans with less than 12-day delinquency, nominated current during the first period, 65.72% remain current in the second period. Furthermore, 16.74% go from 12 to 29 days, 6.09% go from 30 to 45, 2.84% from 46 to 59, 3.54% from 60 to 89, 2.18% from 90 to 119 and 2.89% go to more than 120 days of delinquency during the second period.

	Current	12 - 29 days	30 - 45 days	46 - 59 days	60 - 89 days	90 -119 days	120+ days
Current	65.72%	16.74%	6.09%	2.84%	3.54%	2.18%	2.89%
12 - 29 days	11.99%	26.49%	16.12%	7.93%	7.79%	4.88%	24.80%
30 - 45 days	6.62%	9.19%	18.59%	8.76%	11.11%	5.98%	39.74%
46 - 59 days	3.24%	2.16%	2.70%	4.86%	4.86%	1.62%	80.54%
60 - 89 days	1.35%	1.35%	1.35%	1.35%	3.59%	2.24%	88.79%
90 -119 days	0.00%	0.00%	0.56%	0.00%	0.00%	0.56%	98.88%
120+ days	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Table 6 Current versus Worst Delinquency Comparison for 12-months Period

Of those loans with a delinquency of 12 to 29 days during the first period, 11.99% become current in the second period. In addition, 26.49% go from 12 to 29 days, 16.12% from 30 to 45 days, 7.93% from 46 to 59 days, 7.79% from 60 to 89 days, 4.88% from 90 to 119 days and 24.80% go to more than 120 days of delinquency during the second period.

With regard to loans with a delinquency of 30 to 45 days during the first period, 6.62% become current in the second period. In addition, 9.19% go from 12 to 29 days, 18.59% from 30 to 45 days, 8.76% from 46 to 59 days, 11.11% from 60 to 89 days, 5.98% from 90 to 119 days and 39.74% go to more than 120 days of delinquency during the second period.

Concerning loans with a delinquency period of 46 to 59 days during the first period, only 3.24% become current in the second period. What is more, 2.16% go from 12 to 29 days, 2.70% from 30 to 45 days, 4.86% from 46 to 59 days, 4.86% from 60 to 89 days, 1.62% from 90 to 119 days and 80.54% go to more than 120 days of delinquency during the second period.

Finally, only a small proportion of loans with a delinquency period from 60 to over 120 days during the first period become current in the second period, while the vast majority go to more than 120 days during the second period.

4.2.3. Questionnaire Design

This research complements the above credit data with additional novel information obtained from surveying customers who delay payments. After 12 days of missed payment, these customers are telephoned by a survey company. The customers are asked to complete a questionnaire which investigates the explanations given by respondents for missing payment. In addition, the survey explores psychological traits and levels of financial knowledge which may contribute to understanding the customer's behaviour.

This novel information also assists in establishing the relationship between the time to recovery and the reasons given for non-payment in the survey. The information can improve the prediction of the probability of the borrower being recovered from default and whether LGD is affected by different reasons over time.

A credit limit is given when the card is issued and the customer is free to make as many purchases as they wish (up to the credit limit). These purchases are repaid in fixed monthly instalments and the minimum repayment is the sum of the monthly instalments of each purchase. The customer is considered delinquent after 12 days missing payment when they are approached by the recovery team. However, after 45 days, if debtors have not paid, their debt are sent to collection companies and also the borrowers are considered in default.

The questionnaire was designed (see Figure 3) with a view to conducting a tele-phone survey. The advantages of conducting surveys by telephone are that they are far cheaper and also quicker to administer than other methods and the costs are lower than those incurred for personal interviews.

The lender (who is also a data provider) granted permission to the researcher to approach the customers for the purpose of administering this questionnaire. Thereafter, in August 2013, a pilot was developed and an attempt was made to use a data provider call centre. This, however, proved to be unworkable owing to the sensitive nature of the personal financial information and the inexperience of the staff in this field. Consequently, it was necessary for me to find an alternative organisation that was competent in applying such surveys.

AGP Pesquisas Estatistica, which specializes in surveying such markets was approached to work on this project. AGP is a specialist in achieving both qualitative and quantitative insight through telephone research and has successfully completed research for organizations of all sizes across Brazil.

This company applied the survey based on the premise of a 10 per cent response rate, which means that in order to receive 500 respondents, 5000 customers need to be contacted. Based on figures, a data simple containing 7000 customers (800 up to date, 2,800 delinquents and 3,400 defaulters) were randomly selected from the source population and then it was sent to AGP. The survey proper commenced in May of 2014 and 534 questionnaires had been completed by the end of September 2014, the completion date. Figure A.3 in the appendix shows the answer frequency for each question and its value.

The questionnaire, was produced in English and translated into Portuguese, since the survey was conducted in Brazil, where the company that provides the data is located.

Questions 1 to 11 were intended to elicit information about the respondent's level of knowledge with regard to how banks operate and their responses to default in payment. In addition, the respondent's grasp of basic arithmetic was also tested with a view to gauging their capacity to understand finance. Questions 12 to 21 were introduced in an attempt to ascertain whether or not there is a correlation between risk taking and default. Questions 22 to 24 were included to help provide more information about the respondent's social background.

4.2.4. Distribution for Questionnaire Data

After merging the questionnaire data which to the loan data it was possible to calculate the recovery rates for those customers who were surveyed. This procedure consists in subtracting all payments that occurred after the default date from the balance at the time of default. The data was divided into four groups according to the maximum number of missing payment days at the time of the data extraction. The recovery rate distribution for the questionnaire sample (see Table 7) indicates that recovery rates are widely distributed. Higher intensity is observed at the extremes (0, 1) meaning that for recovery rates equalling zero there was no recovery and for recovery rates equalling one the full amount of the debt was recovered. Figure 7 demonstrated that recovery rate is therefore bounded between zero and one.

Groups	Days missing payment	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Gr_01	>=12 and < 25	88	18.97	88	18.97
Gr_02	>=25 and < 30	70	15.09	158	34.05
Gr_03	>=30 and < 45	95	20.47	253	54.53
Gr_04	>=45	211	45.47	464	100.00

Table 7 Questionnaire Data by Days Missing Payment

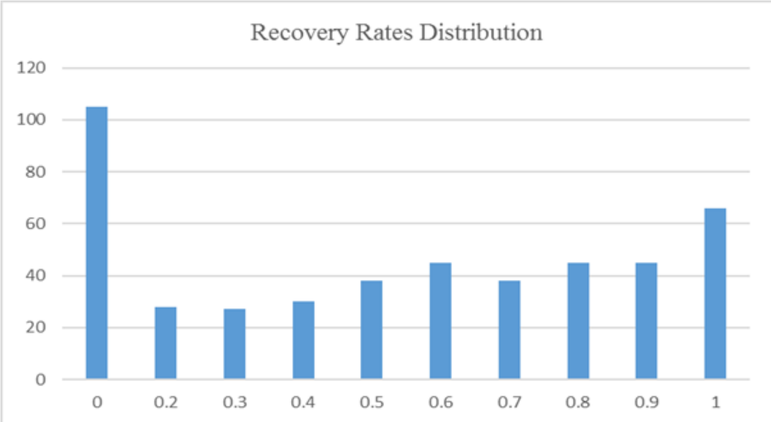


Figure 7 Questionnaire Data RR Distribution

4.2.5. Variables Selection

A common problem in regression is variable selection. There is often a need to select a few from a large number of potential independent variables, perhaps to create the best model. Thus, the main objective of a variable selection procedure is to identify the correct predictor variable which has an important influence on the response variable and could provide robust model prediction.

In this experiment, the variables were selected by testing the predictive power and goodness of fit. The variables were categorised and models were built introducing individual categories on each occasion, and then the p-values of the covariates were assessed for the relevant selection. Thus, those which presented a level of significance of 1%, 5% and 10% were used in the final model.

Another alternative would Stepwise method but it would not be appropriate since this method could exclude variables based on their statistical significance even though these variables would be important to the model due to their economic significance. Moreover, according to Harrell (2001) there a number of problem with Stepwise approach, such as: R^2 values are biased high and the F and X^2 test statistics do not have the claimed distribution.

A number of variables were selected from loan data and from a survey, each question was considered as one covariant and its values are the possible answers to the question. These covariates were categorised as: Loan, Demographic, Reasons for Missing Payment, Risk Taking and Financial knowledge which itself has three subcategories: Accurate Answer, Degree of Accuracy and Score.

Two approaches were considered to cross classify these variables. First, the original value of those variables given above were used in the model. For loan variables, the classes which the data provider used in its credit score model were considered. For the variables that originated from the questionnaire, for those questions that had closed answers, the value of the answers were taken into account. For financial knowledge questions, which had open answers, three possible outcomes were identified: Accurate Answer, which is the specific value of the answer; Degree of Accuracy, which is the distance between the correct answer and the answer given by the respondent; and Score, which was calculated by weighting the answers and adding them together, see appendix 3.

Second, the covariates were coarsely classified and then transformed in two ways: by creating weights of evidence values and secondly, by creating dummy variables, see appendix 4. This was done in order to manage the size of the survey data. Later, the models were built adding each category of variable.

4.3. Methodology

According to Bellotti and Crook (2009) there is a convention in the literature that the LGD models are built in terms of Recovery Rates (RR) rather than LGD directly, where RR = 1 – LGD. This study follows this convention and model LGD in terms of RR, which is written as follow:

$$RR = \frac{sum\ of\ repayment\ made\ over\ a\ period\ t\ following\ default}{outstanding\ balance\ at\ date\ of\ default}$$

The models were built using two approaches: Ordinary Least Square and Zero One Inflated Beta Regression. Also, the variables were transformed.

4.3.1. Ordinary Least Square

The first approach used in this thesis is the Ordinary Least Square (OLS), where the dependent variable is recovery rates and the independent variables refer to borrowers' answers to a questionnaire and loan information on those respondents.

The modes in this study can be written as (see Table 8):

Model	Equation
1	$\mathrm{Y}_i = \alpha + \sum_{N=1}^{N} \beta_N \mathrm{X}_{Ni} + \varepsilon_i$
2	$\mathrm{Y}_i = \alpha + \sum_{N=1}^{N} \beta_N \mathrm{X}_{Ni} + \sum_{N=1}^{N} \delta_N \mathrm{Z}_i + \varepsilon_i$
3	$\mathrm{Y}_i = \alpha + \sum_{N=1}^{N} \beta_N \mathrm{X}_{Ni} + \sum_{N=1}^{N} \delta_N \mathrm{Z}_i + \sum_{N=1}^{N} \gamma_N R_i + \varepsilon_i$

Model	Equation
4	$Y_i = \alpha + \sum_{N=1}^{N}\beta_N X_{Ni} + \sum_{N=1}^{N}\delta_N Z_i + \sum_{N=1}^{N}\gamma_N R_i + \sum_{N=1}^{N}\kappa_N C_i + \varepsilon_i$
5	$Y_i = \alpha + \sum_{N=1}^{N}\beta_N X_{Ni} + \sum_{N=1}^{N}\delta_N Z_i + \sum_{N=1}^{N}\gamma_N R_i + \sum_{N=1}^{N}\kappa_N D_i + \varepsilon_i$
6	$Y_i = \alpha + \sum_{N=1}^{N}\beta_N X_{Ni} + \sum_{N=1}^{N}\delta_N Z_i + \sum_{N=1}^{N}\gamma_N R_i + \sum_{N=1}^{N}\kappa_N E_i + \varepsilon_i$
7	$Y_i = \alpha + \sum_{N=1}^{N}\beta_N X_{Ni} + \sum_{N=1}^{N}\delta_N Z_i + \sum_{N=1}^{N}\gamma_N R_i + \sum_{N=1}^{N}\sigma_N F_i + \varepsilon_i$
8	$Y_i = \alpha + \sum_{N=1}^{N}\beta_N X_{Ni} + \sum_{N=1}^{N}\delta_N Z_i + \sum_{N=1}^{N}\gamma_N R_i + \sum_{N}^{N}\kappa_N D_i + \sum_{N=1}^{N}\sigma_N F_i + \varepsilon_i$

Table 8 OLS Models

Where, Y_i = Recovery Rate of Borrower i.

The exploratory variables are detailed in table 9.

Variables	Description
X	Loan Variables
Z	Demographic Variables (from questionnaire)
R	Reason for Missing Payment (from questionnaire)
C	Financial Knowledge (accurate answer)
D	Financial Knowledge (degree of accuracy)
E	Financial Knowledge (scores)
F	Risk-taking Questions

Table 9 Group of Variables

4.3.2. Zero One Inflated Beta Regression

The second approach used was Zero One Inflated Beta Regression because the beta distribution is very flexible for modelling data that are measured in a continuous scale on the open interval (0,1) since its density has quite different shapes depending on the values of the two parameters that index the distribution (Ferrari and Cribari-Neto 2004)

The beta distribution with parameters μ and ϕ $(0 < \mu < 1$ and $\phi > 0)$, denoted by $\beta(\mu,\phi)$, has density function:

$$f(y;\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(1-\mu)\phi}y^{\mu-1}(1-y)^{(1-\mu)\phi-1}, \quad y\in(0,1),$$

Where $\Gamma(.)$ is the gamma function. If $y\sim\beta(\mu,\phi)$, then $\mathrm{E}(y)=\mu$ and $Var(y)=V(\mu)/(\phi+1)$, where $V(\mu)=\mu(1-\mu)$ denotes the 'variance function'. The parameter ϕ plays the role of a precision parameter in the sense that, for fixed μ, the larger the value of ϕ, the smaller the variance of y. Different values of the parameters generate different shapes of beta density.

In practical applications the data may include zeros/or ones. The beta distribution is not suitable for modelling the data in these situations. The zero-and-one-inflated beta distribution facilitates modelling fractional or proportional data that contains both 0 and 1 (Ospina and Ferrati 2010).

The general idea is to model the response variable as a mixture of Bernoulli and beta distributions, from which the true 0 and 1, and the values between 0 and 1 are generated, respectively. Inflated beta distributions incorporate degenerate probability statements producing a mixture density. For Zero-Inflation, a new parameter $\pi 0$ is added to account for the probability of observations at zero. The subsequent mixture density is:

$$f(\gamma,\pi_0,\mu,\phi)=\begin{cases}\pi_0, & if\ Y=0\\(1-\pi_0)f(\gamma,\pi\_0,\mu,\phi), & if\ 0<Y<1\end{cases} \quad (1)$$

The one inflated methodology follows the same logic as the zero inflated methodology. Here the new parameter π_1 is added to account for the probability of observation at one. The subsequent mixture density is:

$$f(\gamma, \pi_1, \mu, \phi) = \begin{cases} (1-\pi_1) f(\gamma, \mu, \phi), & if\ 0 < Y < 1 \\ \pi_0, & if\ Y = 1 \end{cases} \tag{2}$$

Finally, Zero-One Inflated Beta Regression combines two prior inflated densities into one density:

$$f(\gamma, \pi_0, \pi_1 \mu, \phi) = \begin{cases} \pi_{0,} & if \quad Y = 0 \\ (1-\pi_0)(1-\pi_1)\, f(\gamma, \mu, \phi), & if\ 0 < Y < 1 \\ \pi_{1,} & if \quad Y = 1 \end{cases} \tag{3}$$

The Zero-One Inflated Beta models in this study can be written as:

$\pi_0 = \exp(zeroxb_i)/(1+\exp(zeroxb_i))$

$\pi_1 = \exp(zeroxb_i)/(1+\exp(zeroxb_i))$

$mu = \exp(zeroxb_i)/(1+\exp(zeroxb_i))$

$phi = \exp(d0)$

$w = mu * phi$

$t = phi - mu * phi$ When $rr = 0$ then $Y = \log(\pi_0)$ Thus,

When $rr = 1$ then $Y = \log(\pi_1)$

When $0 < rr < 1$ then

$Y = \lg amma(w+t) - \lg amma(t) + ((w-1) * \log(rr)) + ((t-1) * \log(1-rr)) + \log(1-\pi_0) + \log(1-\pi_1)$

Table 5 shows the equation for each model. Where the exploratory variables are the same variables that were used on the OLS model, please see Table 10.

Model	Equations
1	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N \mathrm{X}_{Ni}$
2	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N \mathrm{X}_{Ni} + \sum_{N=1}^{N} \delta_N \mathrm{Z}_i$

Model	Equations
3	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N X_{Ni} + \sum_{N=1}^{N} \delta_N Z_i + \sum_{N=1}^{N} \gamma_N R_i$
4	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N X_{Ni} + \sum_{N=1}^{N} \delta_N Z_i + \sum_{N=1}^{N} \gamma_N R_i + \sum_{N=1}^{N} \kappa_N C_i$
5	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N X_{Ni} + \sum_{N=1}^{N} \delta_N Z_i + \sum_{N=1}^{N} \gamma_N R_i + \sum_{N=1}^{N} \kappa_N D_i$
6	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N X_{Ni} + \sum_{N=1}^{N} \delta_N Z_i + \sum_{N=1}^{N} \gamma_N R_i + \sum_{N=1}^{N} \kappa_N E_i$
7	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N X_{Ni} + \sum_{N=1}^{N} \delta_N Z_i + \sum_{N=1}^{N} \gamma_N R_i + \sum_{N=1}^{N} \sigma_N F_i$
8	$Zeroxb_i = Onexb_i = Eta_i = \alpha + \sum_{N=1}^{N} \beta_N X_{Ni} + \sum_{N=1}^{N} \delta_N Z_i + \sum_{N=1}^{N} \gamma_N R_i + \sum_{N}^{N} \kappa_N D_i + \sum_{N=1}^{N} \sigma_N F_i$

Table 10 Zero One Inflated Beta Regression Models

4.4. Model Predictive Accuracy Measures

The expected predictive accuracy of the developed models is assessed using four performance measures: the R-squared, the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE).

R-squared can be interpreted as the correlation coefficient between the modelled values to the variability of the original data values. In LGD, it is explicated by variation in the regress ands. R-squared is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2}$$

In statistics, MAE is a measure of how close forecasts or predictors are to the modelled outcomes. It measures accuracy for continuous variables such as LGD. MAE is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left| y_i - \hat{y} \right|$$

MSE measures the average magnitude of the error. RMSE is the root of the mean squared error. Thus, it is the distance, on average, of a data point from the fitted line measure along a vertical line. MSE and RMSE can be expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y} \right)^2 \qquad RMSE = \sqrt{MSE}$$

MAPE (Mean Absolute Percentage Error) is the mean percentage absolute deviation. It generally enunciates the accuracy as percentage, and is defined by the formula:

$$MAPE = \frac{\sum(|A - F|)}{\sum A}$$

4.5. The Models

The objective of the models is to estimate recovery rates following the hypothesis that the reasons for missing payment could help to improve the accuracy of the models. In addition, I will try to establish if there is a relationship between financial knowledge, the propensity to take risks and the ability to pay back a loan.

The models were built by drawing information from the loan database and aforementioned questionnaire. The questionnaire contains four groups of variables namely: demographic, reasons for missing payment, financial knowledge and risk-taking. Eight models were built combining these groups of variables with the variables from the loan data. Two techniques were then applied to each of the models: the first technique was Ordinary Least Square, and second was the Zero One Inflated Beta

Regression. Further models were built using the same techniques, however, the variables were transformed with a view to improving the efficacy of the models. The transformed variables refer to variables categorised into more homogeneous groups in comparison to the original variables. This framework will be explained in the following section. Table 11 shows the summary of the models.

A full analysis, which considers all variables, is presented in appendices 5, 6, 7 and 8, and the models with the significant variables are shown in Tables 12, 13, 16 and 17.

	OLS_O1/ Beta_O1	OLS_O2/ Beta_O2	OLS_O3/ Beta_O3	OLS_O4/ Beta_O4	OLS_O5/ Beta_O5	OLS_O6/ Beta_O6	OLS_O7/ Beta_O7	OLS_O8/ Beta_O8
Original Variables	Loan	Loan, Demographic	Loan, Demographic, Reasons for Missing Payment	Loan, Demographic, Reasons for Missing Payment, Financial Knowledge Accurate Answer	Loan Demographic Reasons for Missing Payment Financial Knowledge Degree of Accuracy	Loan Demographic Reasons for Missing Payment Financial Knowledge Scored Answer	Loan Demographic Reasons for Missing Payment Risk-taking	Loan Demographic Reasons for Missing Payment Financial Knowledge Degree of Accuracy Risk-taking

Note 1 the models, which use transformed variables, have the same groups of variables

Table 11 Summary of Models

4.5.1. Original Variables Models

Eight models were built taking into account the original answers to the questionnaire. The values of the respondents' answers were preserved and incorporated into the models, except for those variables related to financial knowledge, for which three approaches were applied. The first of these approaches was to consider whether or not the respondents' answers were accurate. For those variables, their values were designated 0 for an inaccurate answer and 1 for an accurate answer. The second approach was to calculate the degree of inaccuracy between the correct answer and the incorrect one. Since it was not possible to employ Zero One Inflated Beta Regression using these variables, it was necessary to standardise them. Finally, the third approach consisted in weighting each question according to the degree of difficulty; then a score was created, which in turn was incorporated into the model.

The questionnaire was designed with a view to achieving three outcomes: first, reason for missing payment, second, level of borrower financial knowledge and finally, risk taking propensity. Usually LGD models are built considering borrowers' variables obtained from applications. This study investigates whether or not these new variables from the questionnaire improve the performance of the models.

Models OLS_O1 and Beta_O1 only use loan variables which will work as a benchmark to evaluate the performance of the models based on the questionnaire variables. Models OLS_O2 and Beta_O2 add demographic variables. Apart from the previous variables, Models OLS_O3 and Beta_O3 include reasons for missing payment.

In addition, models OLS_O4, OLS_O5, OLS_O6, Beta_O4, Beta_O5 and Beta_O6 used financial knowledge information measured in different ways. In models OLS_O4 and Beta_O4, the answers were treated as a binary variable: incorrect or correct. OLS_O5 and Beta_O5 took into account the degree of accuracy of the answers. In this case it was the distance between the values of the answer given by the respondent and the value of the correct answer. For example, the correct answer to question 3 ("If the chance of getting a disease is 10%, how many people out of 100 would be expected to get the disease?") is 10. If the respondent gave 50 as an answer the value of the variable would be 40. For OLS_O6 and Beta_O6 models the answers were scored depending on the level of difficulty of the question. Assuming that question 5 is more difficult than question 3 the answer to the former would have a greater weight in calculation of the financial knowledge score according to Table A.2 in the Appendix.

Models OLS_O7 and Beta_O7 use the variables loan, demographic, reason for missing payment and risk-taking. In these models, the variable financial knowledge information was replaced by the risk-taking variable. Models OLS_O8 and Beta_O8 include all the aforementioned types of variables where the financial knowledge information was represented by the degree of accuracy.

4.5.1.1. Results of Original Variables Models

This section presents the model results. Tables 12 and 13 contain statistically significant variables only. The full results of these models are presented in Tables A.6 to A.17 in the appendix.

As shown in Table 8, income (between R$ 600 and R$1200) and ratio income/loan amount (between 0.05 and 0.15) are positively associated with recovery rates in all

OLS models using original variables as defined above. This means that an increase in income is related to a large amount of cash collected. On the other hand, employment length (between 20 and 25 years) and monthly interest rates (4.9%, 5% and 6.8%) have a negative relationship with recovery rates, which indicates that high interest rates reduce the chances of recovery. Regarding employment length, although it is statistically significant, its relationship with recovery rates is weak (10% significance level). Among the demographic variables used model OLS_O2 to OLS_O8, religion (African) showed 10% significance in OLS_O2, 5% for the others and a positive relationship with recovering from default. Despite age (between 31 and 40 years) being significant in model OLS_O1 it does not present significance in the other seven models.

From the third model, reason for missing payment was introduced. Among the reasons for missing payment, unemployment shows statistical significance in models OLS_O3, OLS_O4, OLS_O5 and OLS_O6. As expected it is negatively associated with recovery rates, meaning that those borrowers who become unemployed are less likely to pay back after defaulting.

Regarding the financial knowledge variables which were used in models OLS_O4 to OLS_O8, only degree of accuracy, more precisely, the answer to question 3 presents high significance (1% level) and a negative association with recovering from default. That is, the farther the given answer is from the correct answer the lower the recovery.

Among the 60 risk-taking variables, used only in models OLS_O7 and OLS_08, just question 20 neutral scale, which tests ethical behaviour, shows a positive relation to recovering (10% and 5% respectively).

The results of the Beta models using original variables are shown in Table 12. Employment length (between 20 and 25 years) and monthly loan interest rate (6.1%) are significant in all models and have negative association with recovering. Thus, borrowers who have been employed for between 20 and 25 years are less likely to be recovered in comparison to those who are employed for longer than 30 years (the baseline category). Compared to the baseline interest rate (7%) loans granted at 6.1% present a lower recovery rate.

Gender and ratio income by loan amount (≤ 0.05) are only significant in Beta_O1. The former is positively related to recovery rate while the latter has a negative connection. In other words, female customers tend to recover more than their male counterparts (baseline category) and loans of borrowers with a higher income commitment have less propensity to be recovered.

Other monthly loan interest rate categories which showed statistical significance are 4.9%, 5.0% and 6.1%. The first category (4.9%) is significant only in model Beta_O1, the second (5.0%) is significant in all models except in Beta_O7 and Beta_O8. Finally, interest rate (6.1%) is significant in models Beta_O1, Beta_O4 and Beta_O6. All of them have negative relationship towards recovery rate, meaning that compared to the benchmark rate (7%), the lower the loan interest rate the less successful the collection process would be.

Ratio income by instalment (between 0.5 and 1.0) has a positive relationship with recovery rate in Beta_O1, Beta_O4 and Beta_O6. Another variable that shows positive association with recovery rates is religion (spiritualism). This variable was introduced to the models from Beta_O2 and its significance only appeared in model Beta_O3.

Education level (primary school) is significant only in models Beta_O2 and Beta_O3 and presents negative association with recovery rate. This relationship is related to lower recovery for those who have a lower level of education.

Regarding financial knowledge variables used in models Beta_O6 to Beta_O8, two of them were significant: degree of accuracy (more precisely, the answer to question 6) and question 7, which refers to factors important to borrowers when deciding to take out loans (loan length, interest rate, instalment value). The first one presented high significance (1% level) in Beta_O8 and negative association with recovering from default. The latter is shown to be significant in models Beta_O3 and Beta_O8.

Risk-taking variables were introduced from model Beta_O7 and three of them were significant in Beta_O7 and Beta_O8. Question 16 (answer neutral) is negatively linked to recovery rates while question 19 (answer disapprove) and question 20 (answer neutral) are positively related to recovery rates.

Original variables	OLS_O1		OLS_O2		OLS_O3		OLS_O4		OLS_O5		OLS_O6		OLS_O7		OLS_O8	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	0.758**	0.321	0.8324**	0.3315	1.0424***	0.3441	1.0528***	0.3523	1.059***	0.3492	1.0256***	0.3487				
d_age_2	-0.0969*	0.0568														
d_emp_time_4	-0.1487**	0.0743	-0.1464*	0.0766	-0.1305*	0.0764	-0.1427*	0.0778	-0.1395*	0.0766	-0.1354*	0.0771	-0.1514*	0.079	-0.1601**	0.079
d_sal_1							0.1554*	0.0935								
d_sal_2	0.1446**	0.0619	0.1412**	0.0629	0.1581**	0.0623	0.1603**	0.0637	0.1515**	0.0633	0.1577**	0.0632	0.1802***	0.0657	0.1655**	0.0665
d_lti_1	0.1218*	0.0739	0.1367*	0.0762	0.1542**	0.0755	0.1583**	0.0771			0.1587**	0.0763	0.1338*	0.0793		
d_lti_2	0.1686***	0.0639	0.1863***	0.0659	0.1941***	0.0652	0.1987***	0.067	0.1744***	0.0665	0.2005***	0.0661	0.1703**	0.0688	0.1518**	0.0701
d_lti_3	0.1599***	0.0564	0.1631***	0.0579	0.1794***	0.0571	0.1811***	0.0586	0.1599***	0.0578	0.1837***	0.0578	0.1303**	0.0607	0.1111*	0.0612
d_lti_4	0.138**	0.0607	0.153**	0.0624	0.1799***	0.0615	0.1953***	0.0633	0.17***	0.0627	0.1932***	0.0626	0.1666**	0.0649	0.1666**	0.0664
d_lti_5	0.1094*	0.0588	0.1101*	0.0609	0.1194**	0.0603	0.1179*	0.0621			0.1257**	0.0612	0.1102*	0.0637		
d_int_1	-0.1262**	0.0636	-0.1169*	0.0649	-0.1305**	0.0645	-0.1289*	0.0667	-0.1439**	0.0656	-0.1265*	0.0653	-0.1286*	0.0684	-0.1387**	0.0688
d_int_2	-0.5338***	0.0893	-0.5365***	0.0916	-0.553***	0.0901	-0.551***	0.0924	-0.5575***	0.0906	-0.5429***	0.0913	-0.5179***	0.0951	-0.5148***	0.0952
d_int_4	-0.1093**	0.054	-0.1108**	0.0546	-0.1221**	0.0541	-0.1196**	0.0559	-0.1301**	0.0545	-0.114**	0.0547			-0.0998*	0.057
d_int_5	-0.4249***	0.0601	-0.4306***	0.0608	-0.4389***	0.0601	-0.4432***	0.0617	-0.4561***	0.0608	-0.4363***	0.0606	-0.4343***	0.0618	-0.4516***	0.0622
d_int_6	-0.2324***	0.0756	-0.2344***	0.0764	-0.242***	0.0754	-0.2387***	0.0773	-0.2515***	0.0757	-0.2421***	0.076	-0.2172***	0.0797	-0.2297***	0.0797
d_rel_3													0.1676*	0.0966	0.1697*	0.0975
d_rel_4			0.204*	0.1137	0.2274**	0.1121	0.2264**	0.1143	0.2281**	0.1125	0.2329**	0.1132	0.2544**	0.1217	0.2601**	0.1218
d_rea_2					-0.2076*	0.1151	-0.2011*	0.1176	-0.2073*	0.1159	-0.2105*	0.1167				
sqdist03									-0.0653***	0.0204					-0.0677***	0.0208
d_20_3													0.15*	0.0784	0.1814**	0.0787

Note 1: SE stands for Standard Error

Note 2: ***, **, * indicate 1% ,5% and 10% significance levels, respectively.

Table 12 Original Variables: Significance for OLS Models

Original variables	Beta_O1		Beta_O2		Beta_O3		Beta_O4		Beta_O5		Beta_O6		Beta_O7		Beta_O8	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
d_gen	0.1815*	0.09871														
d_emp_time_4	-0.5217**	0.2453	-0.545**	0.2449	-0.4644*	0.2454	-0.5748**	0.2462	-0.485*	0.2518	-0.6088**	0.2477	-0.5627**	0.2531	-0.5255**	0.2567
d_sal_2_e											0.358*	0.2093	0.375*	0.2096	0.4061*	0.2128
d_int_1	-0.4069*	0.2383														
d_int_2	-1.0025**	0.3984	-0.7784**	0.3912	-0.7075*	0.3866	-0.8643**	0.3993	-0.7929**	0.3943	-0.8763**	0.3874				
d_int_4	-0.4102**	0.1996					-0.3356*	0.197			-0.3527*	0.1909				
d_int_5	-0.6812***	0.2356	-0.5071**	0.2265	-0.4418**	0.2214	-0.6064***	0.228	-0.555**	0.2211	-0.5934***	0.2237	-0.4603**	0.2256	-0.494**	0.2247
d_dti_1	0.5586*	0.3383					0.6114*	0.3563			0.5901*	0.3536				
d_lti_1	-0.4656*	0.2706														
d_edu_1			-0.3322*	0.1833	-0.3224*	0.1796							-0.4315**	0.1915		
d_edu_2													-0.3765**	0.176		
d_rel_2													0.3947**	0.1905		
d_rel_3					0.5229*	0.2881	0.5878**	0.298	0.5814*	0.2969	0.6041**	0.2964	0.5225*	0.2957	0.5316*	0.301
sqdist06															-0.1548***	0.0596
d_cor7_1									-0.2522**	0.1848						
d_cor7_5											-0.04482*	0.2467			-0.511**	0.2583
q_16_3													-0.3648*	0.1961	-0.3642*	0.1988
q_19_2													1.1614***	0.3463	1.0571***	0.3489
q_20_3													0.7571***	0.2461	0.8611***	0.248

Note 1: SE stands for Standard Error

Note 2: ***, **, * indicate 1% ,5% and 10% significance levels, respectively.

Table 13 Original Variables: Significance for Zero One Inflated Beta Models

4.5.1.2. Models Using Original Variables: Performance Comparison

In order to measure the accuracy of each model, four measures were utilized: R-Square, MAE, RMSE and MAPE. Table 14 shows these measures for the eight OLS models based on original variables. According to the four measures used, OLS_O8 perform the best.

	OLS_ O1	OLS O2	OLS_ O3	OLS_ O4	OLS_ O5	OLS_ O6	OLS_ O7	OLS_ O8
R_squared	0.3166	0.3364	0.3709	0.3779	0.3960	0.3749	0.4222	0.4511
MAE	0.2243	0.2212	0.2148	0.2133	0.2106	0.2144	0.2009	0.1974
RMSE	0.2790	0.2800	0.2746	0.2772	0.2731	0.2758	0.2757	0.2732
MAPE	0.377	0.3349	0.3258	0.3236	.3237	0.3255	0.3048	0.3043

Table 14 Performance Measures of OLS Models with Original Variables

Table 15 shows the results concerning the Beta models with original variables. The four measures indicate that Beta_O8 outperforms the other seven models.

	Beta_ O1	Beta_ O2	Beta_ O3	Beta_ O4	Beta_ O5	Beta_ O6	Beta_ O7	Beta_ O8
R_squared	0.0198	0.0141	0.0550	0.0443	0.0539	0.0386	0.0970	0.1028
MAE	0.2459	0.2460	0.2413	0.2389	0.2386	0.2389	0.2264	0.2256
RMSE	0.3198	0.3208	0.3140	0.3158	0.3142	0.3167	0.3070	0.3060
MAPE	0.2584	0.2573	0.2511	0.2447	0.2476	0.2446	0.2283	0.2286

Table 15 Performance Measures of Beta Models with Original Variables

As for the comparison between OLS and Beta models, I focus on the best model for each approach (OLS_O8 and Beta_O8). According to R-squared, MAE and RMSE, the performance of OLS is better than the Beta performance. On the other hand, MAPE suggests the opposite. These results confirm that the use of reason for missing payment, financial knowledge and risk-taking improve the predictive power of recovery rate models.

4.5.2. Transformed Variables Models

Since the number of observations from the questionnaire data was relatively small and as a consequence the frequency of certain variables was similarly low, the variables were grouped into homogeneous categories so as to increase those frequencies. Thus, a further 16 models were built using these new categories that followed the same types of variables in the models based on the original form of the variables. For instance, models OLS_T1 and Beta_T1 correspond with OLS_O1 and Beta_O1 respectively, in terms of variables used.

4.5.2.1. Results of Transformed Variables Models

This section presents the results of the transformed variables models. Tables 16 and 15 contain the statistically significant variables only. The full results of these models are presented in appendix A.

Transformed variables	OLS_T1		OLS_T2		OLS_T3		OLS_T4		OLS_T5		OLS_T6		OLS_T7		OLS_T8	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	0.3743***	0.0807	0.3391***	0.0896	0.3795***	0.0953			0.3125**	0.1503	0.3557***	0.1075	0.6599**	0.2792	0.6542**	0.2957
d_age_1									-0.1026*	0.0563						
d_age_2									-0.1242**	0.0482						
d_sal_3									0.2464***	0.081			0.1011*	0.0601	0.1075*	0.0614
d_prod_1	-0.1078**	0.0491	-0.1099**	0.0493	-0.1087**	0.0491	-0.1056**	0.0502			-0.1036**	0.0495	-0.1126**	0.0505	-0.1136**	0.0511
d_prod_2	-0.1061***	0.0388	-0.1093***	0.039	-0.1111***	0.0388	-0.112***	0.0402	-0.1079**	0.0496	-0.1066***	0.0396	-0.1104***	0.0399	-0.1184***	0.0412
d_dti_1									0.3213***	0.1236						
d_dti_4									0.1372*	0.0747						
d_lti_2	0.135*	0.0697	0.1416**	0.0705	0.1447**	0.0704	0.1491**	0.0721			0.1475**	0.0712	0.1377*	0.0722	0.1513**	0.0737
d_lti_3	0.1195*	0.0617	0.1176*	0.0623	0.1214*	0.0621	0.1232*	0.0636			0.1244**	0.0629	0.1108*	0.0635	0.1139*	0.0648
d_int_1	0.1401***	0.0425	0.1479***	0.0428	0.144***	0.0429	0.1475***	0.0439	0.143**	0.0558	0.1452***	0.0434	0.1485***	0.0437	0.1584***	0.0443
d_int_2	0.1287***	0.0355	0.1325***	0.0357	0.1291***	0.0356	0.1335***	0.0366	0.1005**	0.0443	0.1345***	0.0361	0.141***	0.0363	0.1515***	0.0368
d_rea_1													-0.0683*	0.0404		
d_rea_2					-0.1185**	0.0472	-0.1226**	0.0483			-0.1216**	0.0479	-0.1198**	0.0488	-0.1125**	0.0499
cdist04									0.000303*	0.000165			-0.1323*	0.073		
cdist06															-0.00021*	0.000105
d_20_1													-0.1836**	0.0829	-0.1633**	0.0744
d_20_2															-0.221***	0.0844
d_20_3																

Note 1: SE stands for Standard Error

Note 2: ***, **, * indicate 1% ,5% and 10% significance levels, respectively.

Table 16 Transformed Variables OLS Models Result

Transformed variables	Beta_T1		Beta_T2		Beta_T3		Beta_T4		Beta_T5		Beta_T6		Beta_T7		Beta_T8	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
d_gen_1					0.1615*	0.09587	0.1706*	0.09679			0.1706*	0.09659	0.1642*	0.09547		
d_sal_1	0.2486**	0.1238	0.2897**	0.1217	0.2443**	0.119	0.2953**	0.1207	0.2654**	0.1192	0.2878**	0.1197	0.2244*	0.1186	0.2616**	0.1212
d_int_1	0.2427*	0.1457	0.3001**	0.1453												
d_dti_1							0.5279*	0.3066								
d_lti_2			0.454**	0.2237	0.3753*	0.2176										
d_edu_1			-0.4881***	0.1472	-0.3305**	0.1408	-0.264*	0.1424	-0.2761**	0.1404	-0.2985**	0.1428	-0.3188**	0.1399	-0.2702*	0.1421
d_rea_2					-0.3564**	0.1481	-0.3477**	0.1503	-0.2881*	0.1492	-0.337**	0.1481	-0.2952*	0.1512	-0.2748*	0.1562
score											0.4047*	0.2112				
q_20_1													-0.3982*	0.2062	-0.5314**	0.2118
q_20_2													-0.6464***	0.2326	-0.7688***	0.2391

Note 1: SE stands for Standard Error

Note 2: ***, **, * indicate 1% ,5% and 10% significance levels, respectively.

Table 17 Transformed Variables Zero One Inflated Beta Models Result

As shown in Table 17, loan monthly interest rate ($\leqslant$ 5.8%) and ratio income/instalment (between 0.5 and 1.5) are significant in almost all OLS transformed models except in OLS_T5 in which interest rate ($\leqslant$5.4), and ratio income/instalment (between 0.5 and 1.5) are not significant. They show a positive relationship with recovery rates, in other words, an increase in interest rate increases the probability of recovery

In contrast, the variable product (Sound & Vision, Appliances and Mobile) is negatively associated with recovery rates in all OLS models using transformed variables as defined above. This means that if a loan is taken out for buying these products, it is unlikely to be received back when the borrower defaults.

Once again, reason for missing payment is introduced from the third model. Among the reasons for missing payment, the health problems and unemployment variable show statistical significance in models OLS_T7. However, in model OLS_T3, OLS_T4, OLS_T6 and OLS_T6 only the health problems variable is significant. As expected, in all models these variables are negatively associated with recovery rates, meaning that those borrowers who become unemployed or face health problems are less likely to repay after defaulting than those who face unexpected expenses, which is the benchmark variable in the models.

Financial knowledge variables are used in models OLS_T4 to OLS_T8. Two of these variables are significant: the degree of accuracy in the answer to question 4 and secondly the degree of accuracy in answer to question 6, the former presents significance in OLS_T4 while the latter shows significance in OLS_T8. The former has a positive relationship to recovery rates and the latter is negatively associated with recovering from default. Therefore, because the level of difficulty for question 4 is lower than that for question 6, it could be deduced that borrowers who have less awareness of finance are more likely to be recovered.

Model OLS_T05 is shown to be completely different from the other seven models since the majority of the variables that present significance for this model are not significant in the other models. These variables are: age (<60 years) which has a negative relationship with recovery rates and the ratio income/loan amount (⩽0.05 and between 1.5 and 2.5) that presents positive link to recovery rates.

The results of the Beta models using transformed variables are shown in Table 13. Income (< R$1,000.00) shows significance in all models and its relationship with recovery rates is positive. This indicates that an increase in salary is associated with an increase in default recovering.

Gender presents significance in models Beta_T3, Beta_T4, Beta_T6 and Beta_T7 and a positive association with recovery rates. It could be deduced that female borrowers are more likely to be recovered from default than males.

While ratio income/loan amount (⩽ 0.05) shows significance in model Beta_T4, ratio income\instalment (between 0.5 and 1) is significant in models Beta_T2 and Beta_T3. Both variables have positive association with recovery rates. In this case it is possible to conclude that borrowers who commit less income to loans are more likely to pay back if they become defaulters.

Loan monthly interest rate (⩽5.4) has statistical significance in models Beta_T1 and Beta_T2 only. For these two models this variable is positively associated with recovery rates.

Education level (primary school) is used from model Beta_T2 to Beta_T8 and it is significant in all models. This variable has a negative relationship with recovery rates, meaning that customers who have a low level of education tend to present lower recovery rates.

Reason for missing payment (unexpected expenses in general) shows significance for all models since it is introduced in model Beta_T3. It could be inferred that because it is negatively linked to recovery rates, borrowers who claim to be in debt due to this reason tend not to pay back.

Among the financial knowledge variables, the just score variable shows significance in Beta models. It is important to clarify that this variable is used in model Beta_T5 and presents a positive relationship with recovery rates. This signifies that the greater the borrowers' knowledge of finance the more they are keen to pay back their balances.

Variables from the risk-taking questions are used in Beta_T7 and Beta_T8 and answers to question 20 (strongly disapprove and disapprove) are significant in both models showing negative relationship with recovery rates.

4.5.2.2. Models Using Transformed Variables: Performance Comparison

To assess the accuracy of each model using transformed variables, four measures were considered: R-Squared, MAE, RMSE and MAPE, in a similar way to how they were applied to the models using original variables. Table 18 shows these the measures for the eight OLS models based on transformed variables. According to the four measures used, OLS_T8 also has the best performance.

	OLS_ T1	OLS_ T2	OLS_ T3	OLS_ T4	OLS_ T5	OLS_ T6	OLS_ T7	OLS_ T8
R-squared	0.0944	0.1036	0.1164	0.1257	0.2133	0.1228	0.1540	0.1747
MAE	0.2576	0.2571	0.2529	0.2529	0.2257	0.2522	0.2446	0.2416
RMSE	0.3154	0.3155	0.3140	0.3168	0.2975	0.3150	0.3146	0.3153
MAPE	0.3377	0.3391	0.3331	0.3337	0.3207	0.3317	0.3230	0.3227

Table 18 Performance Measures of OLS Models with Transformed Variables

Table 19 demonstrates the results regarding Beta models for transformed variables. The four measure denote that Beta_T8 performs better than the other seven models.

	Beta_ T1	Beta_ T2	Beta_ T3	Beta_ T4	Beta_ T5	Beta_ T6	Beta_ T7	Beta_ T8
R-squared	-0.062	-0.056	-0.050	-0.054	-0.050	-0.052	-0.033	-0.026
MAE	0.2587	0.2566	0.2558	0.2545	0.2517	0.2539	0.2525	0.2490
RMSE	0.3329	0.3320	0.3312	0.3317	0.3311	0.3313	0.3284	0.3273
MAPE	0.2755	0.2700	0.2690	0.2629	0.2642	0.2626	0.2613	0.2546

Table 19 Performance Measures of Beta Models with Transformed Variables

Models that used transformed variables, show similar behaviour to those models built on Original variable. The three measures: R-squared, MAE and RMSE indicate that the OLS_T8 and Beta_T8 fit the data better, however, according to the MAPE results, model OLS_T7 and Beta_T7 perform better.

4.6. Conclusion

This chapter investigates the reasons behind borrowers delaying payments and how these reasons can assist in estimating recovery rates. Moreover, information on financial knowledge and risk-taking propensity was collected by means of a questionnaire in order to assess its possible relationship with the reasons for delinquency. The information used in the models was collected by surveying customers, both defaulters and non-defaulters, from a personal loan portfolio managed by a Brazilian lender.

32 models were tested using OLS and Zero One Inflated Beta regressions. For each regression method, two types of variables were considered: original (the answers given in the questionnaire) and transformed (variables organised into homogenous classes).

For both techniques and both types of variables, the models including reasons for missing payment, risk-taking propensity and financial knowledge measured by degree of correctness presented the best performance. In comparing OLS and Zero One Inflated Beta regressions, the former was found to have performed better.

Among the reasons for missing payment, the health problems and unemployment variables show statistical significance as expected. In all models, these variables are negatively associated with recovery rates, meaning that those borrowers who become unemployed or face health problems are less likely to repay after defaulting than those who face unexpected expenses, which is the benchmark variable in the models.

Reasons for missing payment were introduced from the third model. Among the reasons for missing payment, the health problems and unemployment variables show statistical significance in models the Ordinary Least Square models. As expected, these variables are negatively associated with recovery rates, meaning that those borrowers who become unemployed or face health problems are less likely to repay after defaulting than those who face unexpected expenses.

For all Beta regression models, reason for missing payment (unexpected expenses in general) shows significance. It could be inferred that because it is negatively linked to recovery rates, borrowers who claim to be in debt due to this reason tend not to be repaid.

Regarding financial knowledge, questions with different level of difficulty are asked in the questionnaire. From the answers for these questions, only two variables regarding to the degree of accuracy showed to be significant: first, the degree of accuracy in the answer to question 4 and secondly the degree of accuracy in answer to question 6, the former presents significance in OLS_T4 model while the latter shows significance in OLS_T8 model. The former has a positive relationship to recovery rates and the latter is negatively associated with recovering from default. Therefore, because the level of difficulty for question 4 is lower than that for question 6, it could be deduced that borrowers who have less awareness of finance are less likely to repay when go to default.

In the Beta regression models, just the score variable is significance and shows a positive relationship with recovery rates. This signifies that the greater the borrowers' knowledge of finance the more they tend to pay back their balances.

Therefore, there is evidence that the new information proposed in this study, namely reasons for missing payment, financial knowledge and risk-taking improve the prediction of recovery rates.

Given that the survey deals with sensitive issues, debtors were not asked directly about their financial circumstances. Thus, the questions were formulated with a view to indirectly inferring the borrowers' situation, their understanding about finance and their willingness in taking risk. As a consequence, the main limitation of this study refers to the fact that one cannot be sure whether or not the respondents either did not understand the questions or did not tell the truth.

Further work can be done regarding to reason for missing payment which can be collected by the in-house collection team when the borrowers are contacted. This procedure would increase considerably the number of observation for modelling purpose. This information then can be used to model recovery rates.

Chapter 5. The Impact of Collection Actions on Recovery Rates Using Retail Loan Level Panel Data

5.1. Introduction

Credit risk assessment plays an import role in the credit risk decisions of financial institutions. Basel II banks are permitted to develop and use their own internal risk ratings. The IRB approach is based on three key parameters used to estimate credit risk: Probability of Default (PD), Exposure at Default (EAD) and Loss Given Default (LGD).

The International Accounting Standards Board (IASB) published the final version of the International Financial Reporting Standards 9 (IFRS 9) in July 2014. The final version of IFRS 9 encompasses the classification, measurement and impairment phases of the IASB's project to replace IAS 39 financial instruments: recognition and measurement.

IFRS 9 replaces IAS 39 and the replacement of accounting standards for provisioning was in response to the financial crisis. The timelines in recognising credit losses and the complexity of multiple impairment models as areas of weakness and in need of consideration were highlighted by the regulators.

Under IAS 39 standards, the 'incurred loss' model delays the recognition of credit losses until there is evidence of a trigger event (e.g. 90 days past due or entry into the operational recoveries environment). As the financial crisis unfolded, it became clear that the existing model allowed the postponement of losses. The main objective of the new impairment model is to recognise expected credit losses at all times and to update the total of credit losses recognised at each reporting date in order to reflect any changes in the credit risk profile of an asset.

Under IAS 39 standards, losses can only be considered that arise from past events and current conditions. The effects of future credit losses cannot be taken into account in the calculation of provisions. The requirements of IFRS 9 broaden the information that can be used when determining credit losses. IFRS 9 is effective for annual periods beginning on or after 1 January 2018. The AIB provides no recommendations or guidance on the modelling methodology that banks should adopt for the calculation of ECL (Expect Credit Loss) (Ramirez 2015).

Loss Given Default (LGD) models predict losses as a proportion of the outstanding loan, in the event a debtor goes into default. The LGD model is one of the components on Expected Credit Loss (ECL). The studies undertaken in LGD mainly investigated different modelling algorithms in order to achieve the most accurate estimation of Recovery Rate (RR) = 1 - LGD. It is known that recovery depends on the debt collection process and the various options or actions that collection departments can take. There is almost no literature that explores the impact of the lender's collection actions on RR/LGD. Therefore, this work investigates the role of different collection actions at the loan level of a retail credit product and estimates LGD using OLS in a Panel Data framework. OLS is chosen because it is largely used in the literature due to its superior performance in estimating LGD models (Bellotti and Crook 2012).

5.2. Recovery Rates across the Collection Process

LGD is an input to the Basel II regulatory capital calculation. Industry models for LGD, particularly for consumer lending portfolios, are often built using Ordinary Least Squares regression or regression trees (Bellotti & Crook (2007), Caselli et al. (2008), Gupton & Stein (2002)).

The regulators, the organisers who have the responsibility for standardising the financial sector, acknowledged that there was a delay in the credit loss recognition on financial instruments because of a weakness in the existing accounting standards.

Therefore, in July 2014 the IASB published the final version of the IFRS 9 Financial Instruments which brings together the classification and measurement, impairment and hedge accounting phases of the IASB's project to replace the IAS 39 financial instruments. This final version must be adopted by all financial companies by January 2018.

IFRS 9 is a forward-looking "expected loss" impairment standard that requires financial companies to provide more timely recognition of expected credit losses (ECL) based on future expectations – as opposed to the current "incurred loss" model. Indeed, the new standard obligates banks to account for ECL on an individual financial instrument level from the moment instruments are first recognized. They must recognize full lifetime ECL on a more timely basis. IFRS 9 effectively demands that accounting statements provide a more accurate view of a bank's financial situation by bringing the impairment methodology used within finance closer to the risk processes employed in expected loss calculations under the Basel regime (Pacter 2014).

5.2.1. Expected Credit Loss

Expected Credit Loss can be defined as losses of interest and principal which occur frequently within the credit environment, since there will always be borrowers who default on their borrowing obligations. ECL is calculated based on three components: Probability of Default (PD), Exposure at Default (EAD) and Loss Given Default (LGD) which are interconnected. It also requires prognoses that should respond well in economic downturns (Botzem 2014).

Under IFRS 9 all assets must be assigned to one of three different categories dependant on their risk profile and arrears status. Depending on the stage allocation, either a 12-month ECL or a Lifetime ECL calculation will be required. A Lifetime ECL is the total amount the bank is expected to lose if the loan is likely to default. The 12-month ECL is a portion of the Lifetime ECL weighted by the probability that the asset will enter default in the following 12 months (Pacter 2014).

These three stages require specific provision methodology to be applied and interest treatment, as demonstrated by Figure 9.

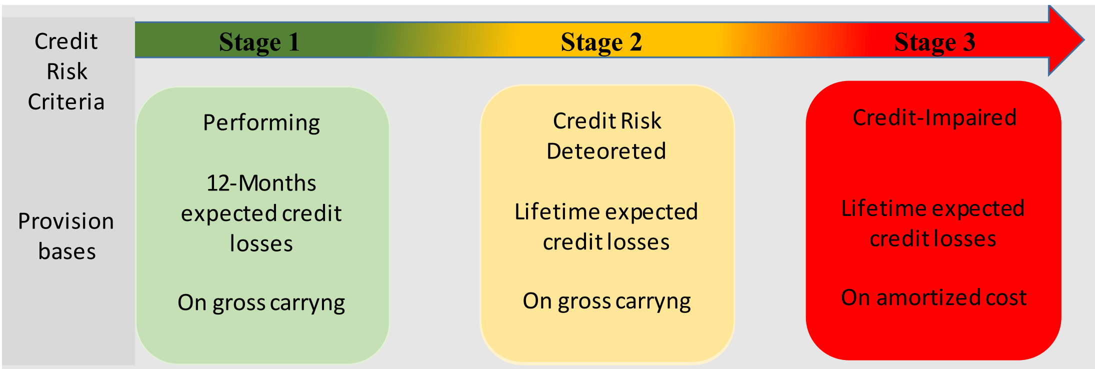


Figure 8 IFRS 9

The lifetime ECL should only be calculated if the credit risk increases significantly from the point of account origination. Therefore an event or trigger must be established. When an account reaches 30 days past due, this could reasonably be defined as having reached Stage 2 classification.

An account can however transition from Stage 2 back to Stage 1 if the credit risk quality for the loan improves to the extent where the risk is similar to accounts classified in Stage 1, however, a probationary period should be established. On the other hand, once an account is classified as Stage 3, account cure is unlikely although possible if all arrears are cleared.

As opposed to the current "incurred loss" model. Specifically, the new standard requires banks to account for ECL on an individual financial instrument level from the moment instruments are first recognised. They must recognise full lifetime ECL on a more timely basis. IFRS 9 effectively demands that accounting statements provide a more accurate view of a bank's financial situation by bringing the impairment methodology used within finance closer to the risk processes employed in expected loss calculations under the Basel regime (Botzem 2014).

Therefore, it is crucial to build accurate models for Probability of Default (PD), Exposure at Default (EAD) and Loss Given Default (LGD). In this context, I build a Panel Data at loan account level, which is crucial for IFRS 9 purposes. The data contains a borrower's characteristics, loan financial information, collection actions over time and the amount of money recovered over time.

Panel Data Modelling lends itself more to the incorporation of time-series variables which may be more powerful predictors than only point-in-time characteristics. This, in turn, also allows integration of macro-economic variables (Baltagi 2008).

5.2.2. Loss Given Default

There are a small number of research papers that explore collection processes in modelling recovery rates. Thomas et al. (2016) built an LGD using the sequence of consecutive payment or non-payment after default. This study applied Markov Chain approaches on portfolio level data to calculate average recovery rates at each stage of the collection process. However, it did not consider the inclusion of economic variables, which, according to the authors, would improve the model's accuracy since economic conditions could affect repayment behaviour.

According to Thomas et al. (2016), there are three main causes of difficulty in modelling LGD: the first is LGD distribution, the second is related to the lender's collection policy and the third is economic effect. Furthermore, it depends on the uncertainty of whether a defaulter will repay, or how much they can afford to repay.

Moreover, Thomas et al. (2012) point out that there are a good number of researches investigating LGD models for corporate loans but relatively few dedicated to unsecured loans. Matuszyk et al. (2010) argued that modelling LGD for unsecured loans is problematic because its outcome is dependent on both lender strategy and debtor affordability. This study is focused on lender collection processes and seeks to define collection strategy.

5.3. Data Description

The purpose of this section is to give an overview of the data provider collection process and how the data was collated to facilitate the development of the model.

5.3.1. Business Understanding

The data provider, which is a finance company from Brazil, has run a store card since 2000. This financial product consists of granting a loan to those customers who buy from its stores and who prefer not to pay in cash. The customer can choose from a range of repayment plans that offer a variety of loan terms and interest rates. The selected interest rate is applied monthly and the customer should pay fixed instalments throughout the contract. To be eligible for the loan, the borrower should present proof of income and residence. At the time of the application, the company records are checked to ascertain that the applicant has never been in debt to Gestao (the data provider). Following this check, the SPC (the credit service protection agency) is consulted in order to guarantee that the customer is not in debt to other financial organisations.

Once the credit is approved, the customer is allowed to conclude the purchase. The first instalment is normally due 30 days after the purchase date and the subsequent payments will be due on the same day each month. If the borrower fails to make a payment within 12 days after the due date, this information will be reported to the collection system. At that point, collection action is initiated in pursuance of the collection policy, with a view to encouraging the borrower to repay what is owed.

5.3.2. Collection Actions

Generally, it can be said, that companies collect the debt mainly in-house and have their own collection departments. However, some companies do use outside agents and from time to time they sell their debt to third parties. Accordingly, the collection process was divided into three phases (See Figure 9): the in-house collection process; the collection process using an agent; and selling the debt.

This illustrates one of the important issues in LGD modelling namely that LGD depends not only on the uncertainty of whether a defaulter will repay, or how much they can afford to repay, but also on the lender's collection policy. For example, the three macro-levels strategies identified above put different bounds on the possible LGD values. When the lender collects the debt in-house they can consider 100% of the payment. As soon as the lenders send to the collection agent, they have to pay a percentage of the money collected to the agent, which varies from 30% to 40%, so it will impact the losses. The last option is to sell off the debt which recovery only 5% of the account balance.

The data provider collection process is based on the length of time during which payments are missed. The first stage of the process involves reminding borrowers of their need to make a payment; this contact is made by telephone after 5 days. The second communication is made by letter and this step is taken after a further period of 12 days without payment having been made. If payment is still not made after a further 25 days, a move is made to the third stage, when the borrowers' details are given to the negative bureau and a letter is sent to inform the borrowers of the action taken. The fourth stage in this process consists of generating data relating to those who have delayed in making payment for a total period of 35 days. This data is then passed on to an external collection company, who assume sole responsibility for collecting the payment. The final stage consists of writing off the contract with the assumption that these loans are lost. The chart below illustrates this process. Figure 1.1 shows the data provider collection strategy.

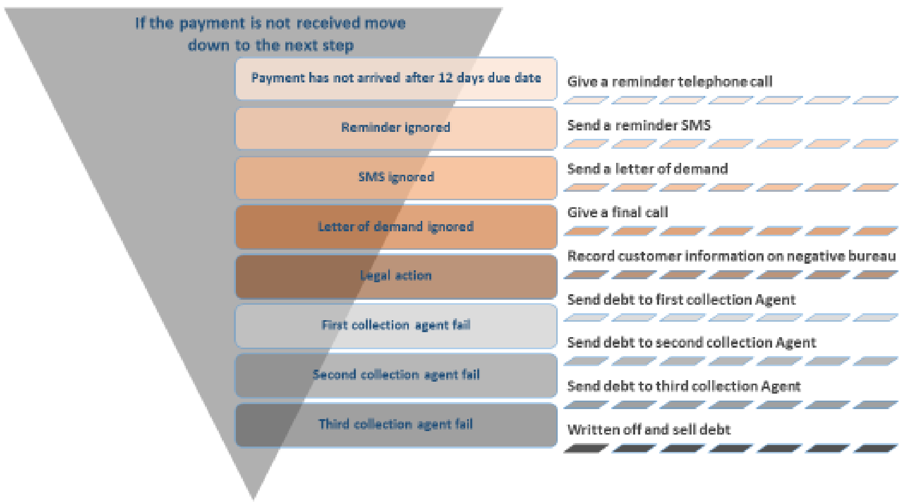


Figure 9 Gestao Collection Strategy

5.3.3. Data Design

Gestao provided the following data which was combined to generate Panel Data to build models to estimate LGD by tracking a borrower's payment following the collection actions:

Loan table: this is a loan table containing information relating to the identification of the customer, the loan account, the date of the application, interests rates, loan amount, loan term, value of the instalment, and the date of the first instalment due.

Payment table: this is a further table with payment details such as the identification of the customer, the loan account, payment value and date of the payment.

Collection table: this table provides dates and codes related to collections actions were also received from the collection system.

Borrower's personal details table: this contains information from the loan application such as date of birth, gender and income.

A daily observation of a loan was recorded from the loan table in order to obtain historic data which would allow the combination of the loan table, the payment table and the collection action table. After amalgamating these tables, it was possible to recognise the daily state of the loan, which is crucial in establishing the correlation between the time of missing payment, its collection actions and payments. This procedure was applied to each loan account where the first date is the date of the loan application and the last date is 31st August 2016. Figure 10 demonstrates how this data was constructed and the steps that were taken in order to obtain the final data that will be used for modelling.

The first step consists of constructing a history table by creating an observation for each loan account for each day from the start of the loan until the end of the observation period of this study, which is 31st August 2016.

The second step involves joining the payment table to the history table by card number, loan account number and date of payment, and then creating new variables. The variables balance and previous_balance initial values are set as the value of the loan_amount variable. The value of balance over time is obtained by subtracting the value of payment from the previous_balance, the following day, if no payment is forthcoming this previous_balance is then equalised to the balance. The variable days_delay is calculated by taking the difference between the history_date and the instalment_due_date. Days_on_book is obtained from the account_open_date and the history_date with the object of ascertaining the age of the loan on the portfolio, which is expressed in days. The overdue_balance is calculated by adding the instalment_value whose payments are overdue.

The account_close_date variable has an initial value of 31dec3500, and will remain until the loan account balance has been paid off, at which time it will be populated with the date of the last payment.

Thirdly, for those accounts which had missed payment and consequently were sent to the collection process, the variables namely collection code and collection date, were incorporated into the history table (see Figure 9). These variables were recorded as binary where 1 means that an action occurred and 0 that no action occurred. Finally, the borrower's personal information was integrated into the final data. Moreover, these variables were cross classified and computed as dummy variables.

After all these proceedings the final data accumulated was enormous and amounted approximately one Teradata. It became very challenging to compute models using this size of data so it was necessary to reduce the number of observations. This issue was addressed by moving the data from daily to weekly. For this, it was necessary to develop a code to recalculate all the variables by the end of each week of the year (see Figure 8).

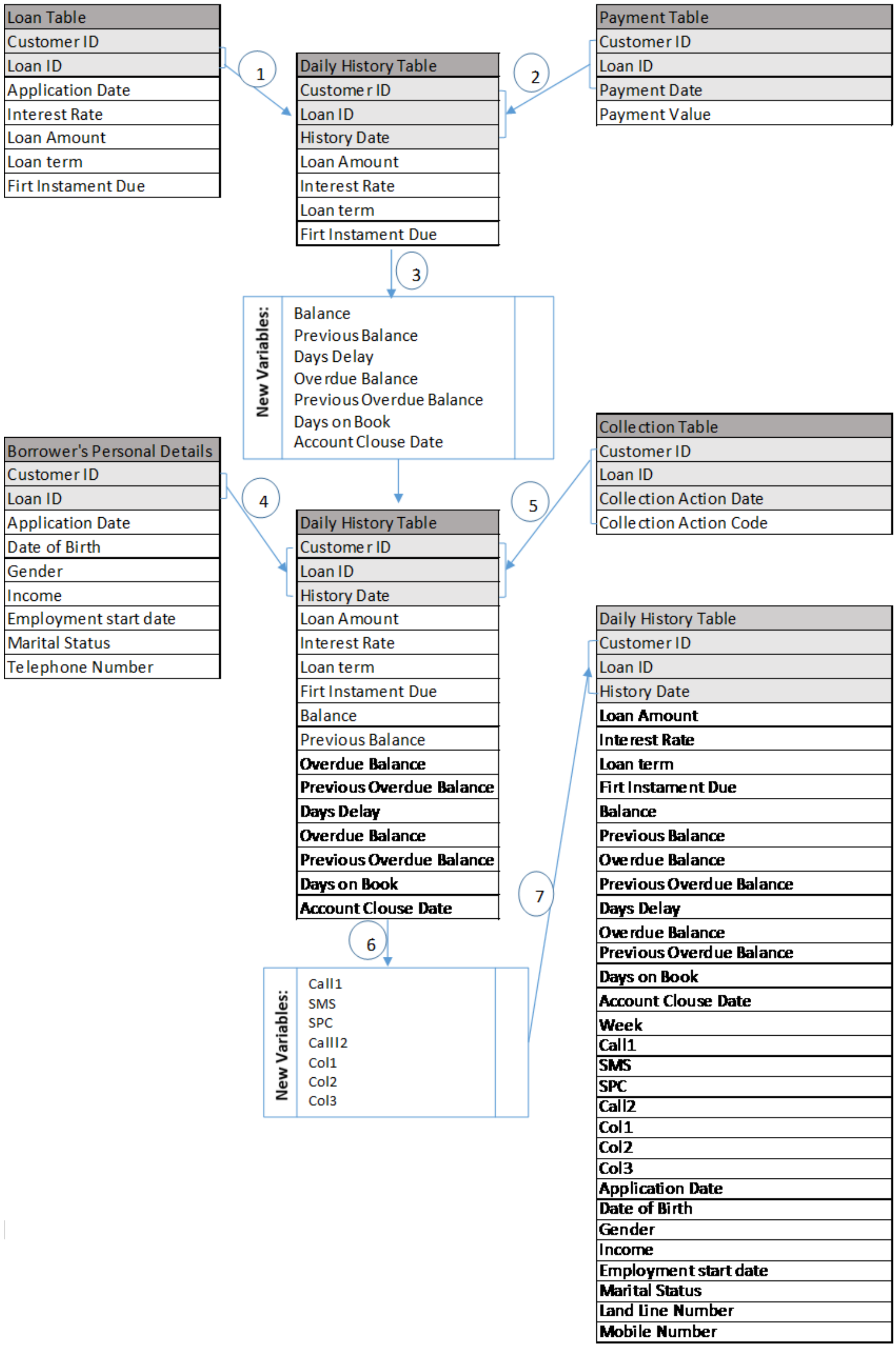


Figure 10 Architecture of the Data

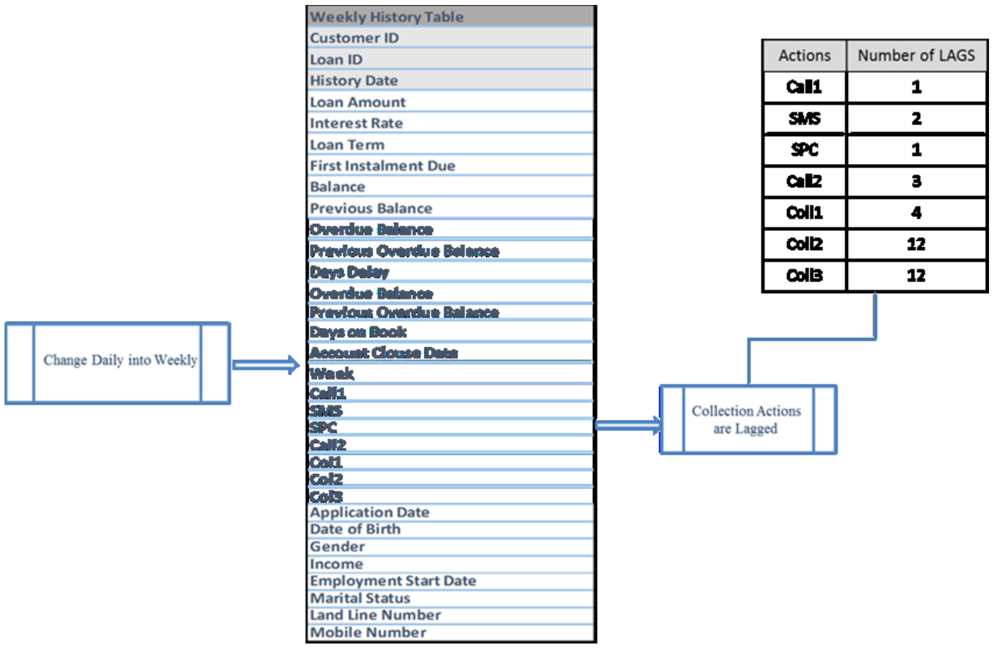


Figure 11 Weekly Lagged Table

The final table has 305 weeks but the number of individuals for each week vary because individuals recover from default at different times, consequently they leave the collection process and are no longer observed. According to the company collection strategy, only those who pay their outstanding balance in full are considered cured, which signifies that once they are sent to the collection system, they will only leave the collection system if they are cured of if there are written off. Table 20 shows the variables which were taken into account for modelling development.

Cross Sectional	Time Series
Age	First Call at 12 DPD
Gender	SMS at 15 DPD
Income	SPC at 25 DPD
Marital Status	Second Call at 30 DPD
Employment Length	First Collect Agency at 60 DPD
Product Category	Second Collect Agency at 90 DPD
Ratio of Instalment/Income	Third Collect Agency at 180 DPD
Ratio of Loan Amount/Income	

Table 20 Covariates Used into the Model

5.3.4. Sample

A sample was extracted for model development. All the accounts which were considered in arrears in 2011 were selected for this work. Figure 12 shows the frequency of the collection action over 2011 and its respective payments. It can be seen that of all the actions, sms and second call collection actions proved to be the most efficient.

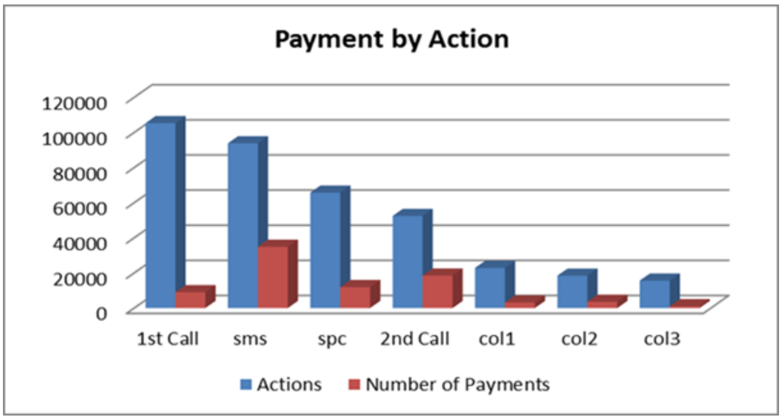


Figure 12 Payment Distribution by Collection Action

Collection actions are divided into in-house and outside house actions. 1st Call, sms, spc and 2nd Call are actions taken by the Gestao collection team, while col1, col2 and col3 are independent collection companies that are contracted to approach customers on behalf of Gestao. These companies charge a value which corresponds to a percentage of the amount recovered and its percentage increases in function of the time in arrears. The information regarding the type of approach applied by these agents is not recorded therefore the action is assumed to be the date on which these accounts are sent to the collection agents.

To measure collection performance, each action is lagged to link each payment to the respective action. The number of lags will depend on the interval between the current action and the next action. These lags are then translated into weeks. Figure 13 shows the payments in-house which were attached to the corresponding lag.

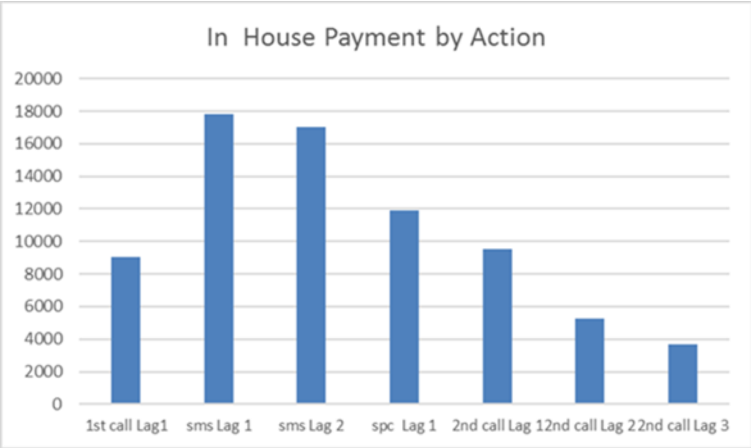


Figure 13 Payment Performance In-house Collection

The time interval for the outside house actions are more spacious therefore the number of corresponding lags is larger. It can be seen from Figure 14, which illustrates the payment distributions that are attributed to outside house actions, the amount paid decreases over time meaning that the longer the time in arrears the greater the difficulty in recovering the money.

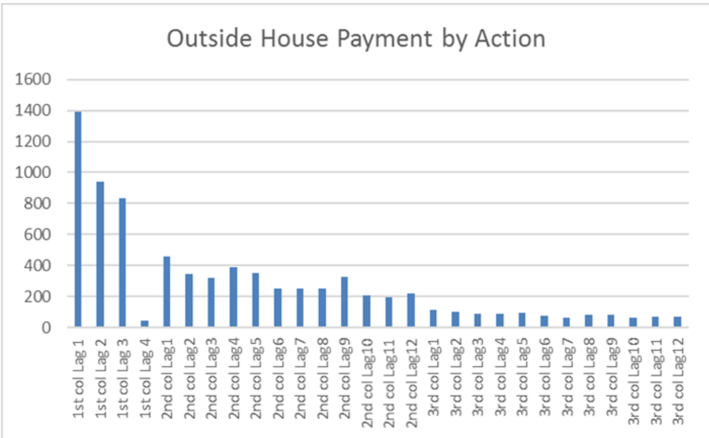


Figure 14 Payment Performance outside House Collection

5.4. Methodology

For this study Panel Data was chosen since it can cope with data which has observations on the same individual over several periods of time (Kennedy 2008). A Panel Data set has multiple entities, each of which has repeated measurements at different time periods. Panel data may have individual effect, time effect, or both, which are analysed by fixed effect and/or random effect models.

As was explain in Chapter 4, the LGD models were built in terms of Recovery Rates (RR) rather than LGD directly.

5.4.1. Models

This study aimed to analyse the effects of collection actions on LGD at account level to comply with IFRS 9 guidance. There are seven collection action in the data provider's collection process. It starts with a call after 12 days due date to give the debtor a remembrance and if the payment is not received it move to the next action until the debt is written off. The waiting time for each action vary from 2 weeks to 12 weeks, it will depend on the collection strategy.

Panel data analysis is a method of studying an exacting subject within multiple sites, periodically observed over a defined period. The combination of time series with cross-sectional data can increase the quality and quantity of data in ways that would be difficult using only one of these two dimensions (Gujarati and Porter 2003). Panel Data analysis can provide a rich and powerful study of a set of people if one is willing to consider both the space and time dimension of the data.

Therefore, to achieve this goal, a Panel Data was built in weekly bases where the payments, which had been made by borrowers in arrears, ware observed over time and linked to the specific lagged action. Thus, RR (Recovery Rates) was estimated by considering the debtor’ s behaviour both across their individual characteristics and by accessing the effect of collection actions over time. Moreover, since the individuals were observed in different period (T≠T for all i), the Panel Data is considered unbalanced.

It is important to highlight that the models were built with a view to measure the impact of the collection actions on the reduction of the outstanding balance and not the impact of approaches since the impact of the collection actions is conditional on what went before and also the possibility of less likelihood to pay.

Four models were built using the Ordinary Least Square (OLS) approach. Two models investigated the impact of collection actions on overdue balances, which is the sum of unpaid instalments, over time. The third explored the effect of loan amount ratio, which is calculated by dividing overdue balance by loan amount on recovery rates. The fourth model considered the effect of ratio, which is the percentage of the outstanding balance that is overdue. Thus, these models sought to establish the best way to estimate recovery rates.

The models can be written as:

$$\mathrm{Y1}_{it} = \alpha_i + \beta_1 D_{it}^{Call1} + \beta_2 D_{it}^{SMS} + \beta_3 D_{it}^{SPC} + \beta_4 D_{it}^{Call2} + \beta_5 D_{it}^{Col1} + \beta_6 D_{it}^{Col2} + \beta_7 D_{it}^{Col3} + \sum_{N=1}^{N} \delta_N \mathrm{X}_{Ni} + \lambda \mathrm{Z}_i + \varepsilon_{i,t}$$

Where for Model 1 and Model 2 $\mathrm{Y1}_{it}$ assumes the changes in overdue balance. For model 3 $\mathrm{Y2}_{it}$ is defined as the ratio calculated by dividing overdue balance by loan amount, while model 4 explores balance ratio which is a ratio obtained by dividing overdue balance by outstanding balance (expressed below). The other regressors are described in Table 21.

$$Y2_{it} = \alpha_i + \beta_1 D_{it}^{Call1} + \beta_2 D_{it}^{SMS} + \beta_3 D_{it}^{SPC} + \beta_4 D_{it}^{Call2} + \beta_5 D_{it}^{Col1} + \beta_6 D_{it}^{Col2} + \beta_7 D_{it}^{Col3} + \sum_{N=1}^{N} \delta_N X_{Ni} + \varepsilon_{i,t}$$

Model Component	Description	Variable Type
$\beta_1 D_{it}^{Call1}$	First Action (telephone call)	Time series
$\beta_2 D_{it}^{SMS}$	Second Action (message)	Time series
$\beta_3 D_{it}^{SPC}$	Third Action (letter)	Time series
$\beta_4 D_{it}^{Call2}$	Forth Action (telephone call)	Time series
$\beta_5 D_{it}^{Col1}$	Fifth Action (1st collecting agent)	Time series
$\beta_6 D_{it}^{Col2}$	Sixth Action (2nd collecting agent)	Time series
$\beta_7 D_{it}^{Col3}$	Seventh Action (3rd collecting agent)	Time series
$\sum_{N=1}^{N} \delta_N X_{Ni}$	Customer's Details	Cross sectional
λZ_i	Outstanding Balance (model 1) Loan Amount (model 2)	Cross sectional
$\varepsilon_{i,t}$	represents the residuals	

Table 21 Exploratory Variables

5.4.2. Models Results

This section presents the model results. Table 22 contains the statistically significant variables pertaining solely to the collection process. The full results of these models are presented in Appendices 29 and 30.

Model 1 shows that the first six actions in the collection process (Call1, sms, spc, Call2, col1, and col2) are negatively associated with overdue balance. On the other hand, the last action (col3) is related to positive changes in the overdue balance.

The results indicate the high efficiency of Call1 and sms given that their coefficients are large compared to the coefficient of the other actions and are statistically significant at the 1% level. It is interesting to note that the strength of the association with sms decreases over time.

SPC, which is the letter sent by the negative bureau (SPC), is also negatively related to overdue balance, though weaker than the previous actions. Call2, the last in-house action, shows a stronger relationship with RR than spc but less so than the two previous actions, as sms and Call2 lose intensity over time.

Although, Col1, the first collection agent, demonstrates a strong link with overdue balance during the first three weeks, this weakens in the fourth week. These results corroborate the patterns illustrated in Figure 14.

Broadly speaking, the relation between col2 and overdue balance is relatively strong over the first three weeks, however, it deteriorates markedly thereafter. What is more, after week nine the direction of this relationship changes from negative to positive.

The trend observed during the last four weeks of col2, col3, is positively associated with overdue balance and the magnitude of this linkage increases with time.

In sum, model 2, which replaces the variable balance with loan amount in comparison to model 1, shows similar overall results. The main exceptions are the size of the coefficients (higher than those in model 1) and a small number of insignificant lags for col2.

With reference to model 3, the dependent variable of which is loan amount ratio (overdue balance divided by loan amount), the results show that the first five actions are negatively connected to loan amount ratio, however, for all col3 lags and most of col2, excluding lag 2 and lag 3, this relationship changes and the association between these variables and the dependent variable changes to positive.

Regarding model 4, whose dependent variable is overdue balance divided by outstanding balance, only variables Call1 and sms have positive association with the dependent variable. The other variables and their respective lags show positive relationship with balance ratio. Furthermore, the coefficients present a clear rising trend from col2 lag 3 to the last lag of col3.

As shown in Table 23, virtually all control variables are statistically significant in the four models.

	Model 1		Model 2		Model 3		Model 4	
Dependent Variable	Overdue balance		Overdue balance		Overdue balance/Loan amount		Overdue balance /Balance	
Collection Actions	Estimate	t	Estimate	t	Estimate	t	Estimate	t
l1_call1	-256.1588 ***	-116.74	-359.5600 ***	-113.74	-0.235292 ***	-147.12	-0.185515 ***	-17.88
l1_sms	-260.9699 ***	-162.04	-367.5741 ***	-158.45	-0.237871 ***	-202.68	-0.185714 ***	-24.39
l2_sms	-185.8208 ***	-80.08	-285.1386 ***	-86.69	-0.172199 ***	-103.48	-0.092479 ***	-8.57
l1_spc	-57.53807 ***	-20.92	-108.8432 ***	-27.47	-0.056049 ***	-27.96	0.013353	1.03
l1_call2	-197.3801 ***	-93.23	-316.1666 ***	-103.66	-0.178800 ***	-115.87	-0.030128 ***	-3.01
l2_call2	-179.2883 ***	-79.55	-291.5136 ***	-89.78	-0.157177 ***	-95.68	0.009174	0.86
l3_call2	-136.9423 ***	-58.22	-229.5463 ***	-67.74	-0.113123 ***	-65.98	0.044117 ***	3.97
l1_col1	-159.551 ***	-52.22	-209.9550 ***	-47.69	-0.093368 ***	-41.92	0.091311 ***	6.32
l2_col1	-164.6274 ***	-53.22	-212.8397 ***	-47.75	-0.094261 ***	-41.80	0.091289 ***	6.24
l3_col1	-165.6753 ***	-53.09	-211.5525 ***	-47.04	-0.092873 ***	-40.82	0.086616 ***	5.87
l4_col1	-113.1542 ***	-35.96	-157.3697 ***	-34.70	-0.051315 ***	-22.37	0.145611 ***	9.78
l1_col2	-105.4304 ***	-31.22	-121.4622 ***	-24.96	0.028189 ***	-11.45	0.156279 ***	9.79
l2_col2	-110.4063 ***	-32.52	-124.8195 ***	-25.52	-0.029901 ***	-12.08	0.153957 ***	9.59
l3_col2	-113.7493 ***	-33.38	-127.2828 ***	-25.92	-0.031372 ***	-12.63	0.148911 ***	9.24
l4_col2	-62.76525 ***	-18.34	-76.7450 ***	-15.56	0.005414 ***	2.17	0.199055 ***	12.30
l5_col2	-40.14955 ***	-11.68	-53.0336 ***	-10.70	0.021274 ***	8.49	0.211601 ***	13.01
l6_col2	-42.72701 ***	-12.39	-53.9376 ***	-10.85	0.020549 ***	8.17	0.205215 ***	12.58
l7_col2	-45.69565 ***	-13.21	-54.3547 ***	-10.90	0.019641 ***	7.79	0.201565 ***	12.32
l8_col2	-23.03841 ***	-6.64	-31.4952 ***	-6.29	0.035928 ***	14.19	0.221493 ***	13.49
l9_col2	15.42757 ***	4.42	7.0828	1.41	0.062719 ***	24.66	0.257530 ***	15.61
l10_col2	12.58423 ***	3.59	7.4236	1.47	0.062754 ***	24.56	0.251635 ***	15.19
l11_col2	9.126362 ***	2.60	4.4572	0.88	0.061636 ***	24.07	0.248535 ***	14.96
l12_col2	13.61649 ***	3.86	10.2389 **	2.02	0.064925 ***	25.26	0.248307 ***	14.90
l1_col3	56.05598 ***	15.17	79.3606 ***	14.91	0.112344 ***	41.71	0.302476 ***	17.31
l2_col3	52.48934 ***	14.16	77.7802 ***	14.56	0.111145 ***	41.12	0.298301 ***	17.02
l3_col3	48.45431 ***	13.02	75.60402 ***	14.10	0.110269 ***	40.66	0.294775 ***	16.76
l4_col3	89.29336 ***	23.93	117.5694 ***	21.86	0.135831 ***	49.93	0.322646 ***	18.28
l5_col3	102.8579 ***	27.48	133.1309 ***	24.68	0.144856 ***	53.08	0.331166 ***	18.71
l6_col3	99.08024 ***	26.4	130.1663 ***	24.07	0.143409 ***	52.41	0.326855 ***	18.42
l7_col3	95.66453 ***	25.43	128.7635 ***	23.75	0.142269 ***	51.87	0.323981 ***	18.21
l8_col3	115.416 ***	30.61	148.9223 ***	27.41	0.153589 ***	55.88	0.334425 ***	18.76
l9_col3	144.1119 ***	38.12	178.7295 ***	32.81	0.170441 ***	61.84	0.352283 ***	19.71
l10_col3	140.5182 ***	37.08	175.8913 ***	32.21	0.168662 ***	61.05	0.347953 ***	19.42
l11_col3	136.79.52 ***	36.02	172.9285 ***	31.60	0.166741 ***	60.22	0.343585 ***	19.13
l12_col3	140.4618 ***	36.89	177.5205 ***	32.35	0.169016 ***	60.89	0.344270 ***	19.12

Note 1: Std. Err. stands for Standa

Note 2: ***, **, * indicate 1%, 5% and 10% significance level,

Table 22 Collection Action Variables: Significance for OLS Models

	Model 1		Model 2		Model 3		Model 4	
Dependent Variable	Overdue balance		Overdue balance		Overdue balance/Loan amount		Overdue balance /Balance	
Control Variables	Estimate	t	Estimate	t	Estimate	t	Estimate	t
intercept	-463.6294 ***	-62.57	-389.3774 ***	-35.57	0.224390 ***	42.06	0.076150 **	2.20
d_gen_1	0.9121282	1.80	-1.0993	-1.51	-0.008860 ***	-24.02	0.055993 ***	23.41
d_age_1	41.64924 ***	43.17	159.8126 ***	115.12	0.118503 ***	168.73	0.081259 ***	17.84
d_age_2	42.97944 ***	47.33	116.1115 ***	127.23	0.111032 ***	168.10	0.134584 ***	31.41
d_age_3	25.7202 ***	28.15	94.7893 ***	72.05	0.062344 ***	93.69	0.051208 ***	11.86
d_age_4	13.63803 ***	14.02	55.1911 ***	39.39	0.042126 ***	59.44	0.022593 ***	4.91
d_sal_1	6.512973 ***	2.88	-168.6496 ***	-48.99	-0.059477 ***	-37.61	-0.044038 ***	-4.29
d_sal_2	6.078524 ***	2.77	-163.3286 ***	-49.88	-0.070415 ***	-45.31	-0.038345 ***	-3.80
d_sal_3	13.72961 ***	6.27	-94.9097 ***	-29.57	-0.039871 ***	-25.32	0.106401 ***	10.42
d_sal_4	7.248032 ***	3.04	-87.1616 ***	-25.18	-0.026030 ***	-15.07	-0.010829	-0.82
d_sal_5	11.77858 ***	4.70	-65.1471 ***	-17.97	-0.025716 ***	-14.11	-0.010829	-0.92
d_sal_6	8.210791 ***	2.77	-11.1957 ***	-2.61	0.003583 *	1.66	0.197025 ***	14.07
d_inter_1	-43.95138 ***	-62.39	-109.2726 ***	-107.75	-0.065620 ***	-127.90	-0.071223 ***	-21.40
d_inter_2	-23.9792 ***	-32.03	-49.1204 ***	-45.52	-0.037934 ***	-69.51	-0.057684 ***	-16.29
d_inter_3	-12.34821 ***	-16.40	-34.9173 ***	-32.18	-0.031230 ***	-56.90	-0.051132 ***	-14.36
d_emp_1	25.18909 ***	13.51	79.3555 ***	29.55	0.039411 ***	29.01	0.032101 ***	3.64
d_emp_2	14.09373 ***	6.52	48.2902 ***	15.51	0.012277 ***	7.79	-0.154688	-1.51
d_emp_3	8.451 ***	-32.52	49.8974 ***	17.89	0.018677 ***	13.24	0.016517 *	1.80
d_emp_4	1.620075	0.72	27.1767 ***	8.38	0.005352 ***	3.26	-0.016242	-1.53
d_emp_5	-8.122247 ***	-3.43	24.2855 ***	7.11	0.021985 ***	12.72	-0.007367	-0.66
d_prod-1	13.46657 ***	20.02	57.0732 ***	58.91	0.045278 ***	92.38	-0.030800 ***	-9.69
d_prod-2	-7.547975 ***	-10.10	-23.8786 ***	-22.18	0.002205 ***	4.05	-0.055150 ***	-15.61
d_prod-3	15.02462 ***	21.73	49.4895 ***	49.70	0.019592 ***	38.89	-0.035560 ***	-10.88
d_dti_1	-157.733 ***	49.49	-132.6172 ***	-64.68	-0.140752 ***	-136.95	-0.287512 ***	-43.13
d_dti_2	-133.2521 ***	-132.20	-89.7592 ***	-61.66	-0.072294 ***	-98.40	-0.217662 ***	-45.67
d_dti_3	-85.63341 ***	-98.93	-46.5119 ***	-37.19	-0.028286 ***	-44.83	-0.133842 ***	-32.70
d_dti_4	-41.37861 ***	-48.97	-2.3748 ***	-1.95	-0.004282 ***	-6.95	0.047043 ***	11.78
d_lti_1	228.8756 ***	163.66	95.4013 ***	40.92	0.120792 ***	132.65	0.316919 ***	53.65
d_lti_2	185.0189 ***	168.91	79.8201 **	43.48	0.083361 ***	117.44	0.235693 ***	51.19
d_lti_3	152.0052 ***	150.58	67.6280 ***	41.48	0.058146 ***	85.87	0.260373 ***	59.28
d_lti_4	109.2793 ***	130.75	57.0230 ***	43.95	0.038370 ***	66.18	0.075574 ***	20.09
d_mst_1	45.10767 ***	31.64	149.3763 ***	14.10	0.068203 ***	65.64	0.061498 ***	9.12
d_landline	5.608163	0.82	16.10412	1.64	-0.018977 ***	-3.81	0.037446	1.16
d_mobile	-89.65093 ***	-92.3	-162.826 ***	-116.43	-0.095545 ***	-136.49	-0.073384 ***	-15.99
week	3.065381 ***	841.66	2.812756 ***	535.97	0.001854 ***	698.40	0.003798 ***	220.54
balance	0.7517179 ***	2197.68						
loan_amount			0.47277 ***	843.00				

Note 1: Std. Err. stands for Standard
Note 2: ***, **, * indicate 1%, 5% and 10% significance level, re
Note 3: The difference among the models refer to different control variables used (see Table 17)

Table 23 Control Variables Significance Level

Measure	Model 1	Model 2	Model 3	Model 4
R-squared	0.7621	0.5060	0.2709	0.0238
Adj R-squared	0.7621	0.5060	0.2709	0.0237
Root MSE	415.81	599.18	0.030314	1.9663
MAE	297.5692210	474.9476184	0.2667748	0.2921785
MSE	466.8474451	650.0159292	0.3142198	0.3100114

Table 24 Models Comparison

Of all the collection actions, those most strongly associated with the highest reduction in overdue balance are Call1 and sms. This could be explained by the fact that these two collection actions aim to remind borrowers about the delay in their payment.

Since delays in this period may be due to financial mismanagement (e.g. borrowers can be delinquent for trivial reasons such as administrative error or forgetfulness (Finlay 2010)), when contacted by the collection team these customers tend to make their payment at once.

The spc action takes place 25 days after the payment is overdue. The customer's information is than sent to the negative bureau (SPC) which sends a letter warning the customer that if payment is not made, their information will be recorded on its database[1].

This collection action show a weaker link among the in-house actions. This is possibly explained by the fact that the borrowers have 10 days to repay their debts before being recorded. Nonetheless, 5 days after spc action, Call2 (a reminder about SPC) is made.

Those who can afford the repayment tend to pay back after Call2 and not just after the spc action. This explains the larger coefficient in magnitude for Call2 in comparison to spc action.

[1] In Brazil, all finance institutions access the SPC (Credit System Protection) database at the time of evaluating credit applications.

The first outside action, col1, seems to perform well since its coefficients are statistically significant and relatively outsized. Two reasons would explain this performance: first, the collection agent is more specialised in approaching debtors in early delinquency, second, debtors are possibly still trying to cope with their financial problems.

Although col2 are also negative their size falls drastically vis-a-vis col1. The likely cause of this is that as time elapses, other instalments become past due, resulting in the increase of the overdue balance and making it harder for the debt to be recovered. This situation is even more evident in col3, as can be seen by its crescent positive coefficient.

5.4.3. Model Validation

The models were validated using a holdout sample which contains loans accounts opened in 2012 which missed payments for more than 12 days during their lifecycle. Table 24 shows that model 1 is the best fit for estimating recovery rates: 76% adjusted R-squared and the one which presents the lower values for MAE and MSE.

5.5. Conclusion

This chapter investigates whether or not collection actions help in estimating recovery rates, which can be utilised as an input of Expected Credit Loss (ECL) calculation in the light of IFRS 9 methodology. I contribute to the literature by incorporating new information on RR/LGD models by combining information on obligors, their collection actions and their payments.

According to the analyses, collection actions are strongly associated with recovery rates in that models which take this information into account would assist financial institutions in estimating LGD and thereby meet IFRS 9 requirements.

Broadly speaking, the efficiency of collection actions diminishes over time. Whereas the actions in the early stages of the collection process are more efficient, in the later stages they seem to be less effective.

Therefore, these results can also support managers when defining their collection strategy.

Further work is needed to incorporate into the model Macroeconomic Variables, which have long-run relationship to personal loan, for stress testing hypothetical changes in the economy.

Chapter 6. Conclusion

6.1. Introduction

Loss Given Default is an important measure of credit loss used by financial institutions to compute risk within credit portfolios, expected loss on individual loans and capital requirements. The Basel II Capital Accord gives banks the opportunity to calculate their own estimates of Loss Given Default.

To address the specific challenges that IFRS 9 poses for retail portfolios, this thesis investigates new information which can help in modelling LGD, which are developed in terms of Recovery Rates (RR) rather than LGD directly, where RR = 1 – LGD and in accordance to IFRS 9 guidance, in three ways: by considering whether or not there is a short and/or a long-run relationship between the delinquency rates in Brazilian consumer personal credit and macroeconomic variables in Chapter 3, secondly, by investigating reasons for missing payments and establishes whether or not these reasons are related to a customer's propensity to risk and the extent of their financial knowledge in Chapter 4, and thirdly, by analysing the impact of collection actions in estimating Recovery Rates, in Chapter 5.

6.2. Contributions

This thesis contributes to the literature by testing new information to estimate loss given default using data from a Brazilian lender and Brazil's economy. Such novel information refers to reasons for missing payment, financial knowledge, risk-taking propensity and collection actions. Furthermore, it investigates a possible relationship between delinquency rates and macroeconomic variables which could be related to recovery rates.

Furthermore, in Chapter 5, LGD was modelled at the account level since the literature in field usually addresses LGD at the portfolio level. This contribution is important to the compliance to IFRS 9 methodology which has been required by regulators from January 2018.

6.3. Findings

Chapter 3 applies a Vector Error Correction Model (VECM) to explore a potential causal link between macroeconomic variables and both aggregate delinquency rates up to and above 90 days. The results suggest that the dependent variable delinquency rates up to 90 days cannot be analysed by VECM since that variable is stationary at (0) level.

As for the long-run relationship, all the independent variables included into the model affect the delinquency rate of more than 90 days. In the short term, three variables are connected to delinquency rate over 90 days: lag 3 of the own dependent variable (i.e. delinquency rates of more than 90 days), lag 3 of unemployment rate and lag 1 of personal loan portfolio outstanding balance.

Even though the model implies long-run relationship for aggregate delinquency rates above 90 days, the values of Adjusted R^2 suggest a low explanatory power of the independent variables.

In Chapter 4, recovery rates were modelled using, besides borrowers' characteristics, customers' reasons for missing payment, individuals' financial knowledge and their propensity to take risk. This information was gathered through a survey answered by the data provider's borrowers.

Chapter 5 explores the impact of collection process on recovery rates and consequently on loss given default, which is one of the parameters used to calculate Expected Credit Loss (ECL) in the context of bank regulation. Combining borrowers' personal information, loan financial details and payments from the collection process was the main contribution of this research.

The findings suggest that there is a strong relationship between the collection actions and recovery rates. This means that financial institutions would be advised to take into consideration information about collection actions in models aimed at estimating LGD in line with the new IFRS 9 methodology. In addition, these results could assist the definition of collection strategies by lenders.

Using the novel information mentioned above, 32 models were built based on OLS and Zero One Inflated Beta regressions. For both approaches, the models including reasons for missing payment, risk-taking propensity and financial knowledge measured by degree of correctness performed best. Regarding the statistical technique, OLS was superior to Zero One Inflated Beta.

The research reveals that the new information proposed here has predictive power on recovery rates. This new information that was investigated in this thesis can be used together in a LGD model by combining reasons for missing payment, collection actions and macroeconomic variables which are co-integrated to aggregate delinquency.

6.4. Limitations and Further Work

This research has certain limitations. First, in Chapter 3, all macroeconomic series related to the credit consumer market were discontinued in December 2012 by the Brazilian Central Bank. It was therefore impossible to include more recent data into the analysis. Second, in Chapter 4, given that the survey deals with sensitive issues, debtors were not asked directly about their financial circumstances. Thus, the questions were formulated with a view to indirectly inferring the borrowers' situation, their understanding about finance and their willingness in taking risk. As a consequence, one cannot be sure whether or not the respondent either did not understand the questions or did not tell the truth. Third, regarding the outside house collection process depicted in Chapter 5, it was assumed that the moment that the debtors' details were sent to the collection company the debtors were contacted by the agent. In the models the date when information was sent to the agent was considered the date of the effective action although such action may not have happened.

Suggestions for further studies in this area would refer to collecting the reasons why borrowers miss payment by the time they are first contacted by the collection team. This would result in more data available and it could give a better picture related to the customers' financial situation. Additionally, the methodologies in this research should be applied using data from other countries and other types of products.

References

Ali, Asghar, and Kevin Daly. 2010. "Macroeconomic determinants of credit risk: Recent evidence from a cross country study." Review of. *International Review of Financial Analysis* 19 (3):165-71.

Altman, Edward I, Brooks Brady, Andrea Resti, and Andrea Sironi. 2005. "The link between default and recovery rates: Theory, empirical evidence, and implications." Review of. *The Journal of Business* 78 (6):2203-28.

Altman, Edward I, and Egon A Kalotay. 2014. "Ultimate recovery mixtures." Review of. *Journal of Banking & Finance* 40:116-29.

Araten, Michel, Michael Jacobs, and Peeyush Varshney. 2004. "Measuring LGD on commercial loans: an 18-year internal study." Review of. *RMA JOURNAL* 86 (8):96-103.

Asarnow, Elliot, and David Edwards. 1995. "Measuring loss on defaulted bank loans: A 24-year study." Review of. *The Journal of Commercial Lending* 77 (7):11-23.

Baldwin, Timothy T, and J Kevin Ford. 1988. "Transfer of training: A review and directions for future research." Review of. *Personnel psychology* 41 (1):63-105.

Baltagi, Badi. 2008. *Econometric analysis of panel data*: John Wiley & Sons.

Bangia, Anil, Francis X Diebold, André Kronimus, Christian Schagen, and Til Schuermann. 2002. "Ratings migration and the business cycle, with application to credit portfolio stress testing." Review of. *Journal of Banking & Finance* 26 (2):445-74.

Bastos, João A. 2010. "Forecasting bank loans loss-given-default." Review of. *Journal of Banking & Finance* 34 (10):2510-7.

———. 2014. "Ensemble predictions of recovery rates." Review of. *Journal of Financial Services Research* 46 (2):177-93.

Bellotti, Tony, and Jonathan Crook. 2009. "Credit scoring with macroeconomic variables using survival analysis." Review of. *Journal of the Operational Research Society* 60 (12):1699-707.

———. 2012. "Loss given default models incorporating macroeconomic variables for credit cards." Review of. *International Journal of Forecasting* 28 (1):171-82.

Bennett, Rosalind L, Eva Catarineu, and Gregorio Moral. 2005. "Loss given default validation." Review of. *Basel Committee on Banking Supervision, Studies on the Validation of Internal Rating Systems, Working Paper* (14):60-93.

Berthoud, Richard, and Elaine Kempson. 1992. *Credit and debt: the PSI report*: Policy Studies Institute London.

Black, Fischer, and John C Cox. 1976. "Valuing corporate securities: Some effects of bond indenture provisions." Review of. *The Journal of finance* 31 (2):351-67.

Blais, Ann-Renee, and Elke U Weber. 2009. "The domain-specific risk taking scale for adult populations: Item selection and preliminary psychometric properties." In.: DTIC Document.

Botzem, Sebastian. 2014. "Transnational standard setting in accounting: Organizing expertise-based self-regulation in times of crises." Review of. *Accounting, Auditing & Accountability Journal* 27 (6):933-55.

Bucher, Jay L. 2012. *The metrology handbook*: ASQ Quality Press.

Calabrese, Raffaella. 2012. "Estimating bank loans loss given default by generalized additive models." Review of. *UCD Geary Institute Discussion Paper Series, WP2012/24*.

Cameron, Samuel, and D Golby. 1991. "Personal debt crises: an economic approach." Review of. *International Review of Applied Economics* 5 (3):310-24.

Capon, Noel, Gavan J Fitzsimons, and Russ Alan Prince. 1996. "An individual level analysis of the mutual fund investment decision." Review of. *Journal of Financial Services Research* 10 (1):59-82.

Caselli, Stefano, Stefano Gatti, and Francesca Querci. 2008. "The sensitivity of the loss given default rate to systematic risk: new empirical evidence on bank loans." Review of. *Journal of Financial Services Research* 34 (1):1-34.

Chalupka, Radovan, and Juraj Kopecsni. 2008. "Modelling bank loan LGD of corporate and SME segments: A case study." In.: IES Working Paper.

Collin‐Dufresne, Pierre, and Robert S Goldstein. 2001. "Do credit spreads reflect stationary leverage ratios?" Review of. *The Journal of finance* 56 (5):1929-57.

Committee, Basel. 2005. "An Explanatory Note on the Basel II IRB Risk Weight Functions." Review of. *Bank for International Settlements*.

Cowan, Adrian M, and Charles D Cowan. 2004. "Default correlation: An empirical investigation of a subprime lender." Review of. *Journal of Banking & Finance* 28 (4):753-71.

Crook, Jonathan, and John Banasik. 2012. "Forecasting and explaining aggregate consumer credit delinquency behaviour." Review of. *International Journal of Forecasting* 28 (1):145-60.

Davies, Emma, and Stephen EG Lea. 1995. "Student attitudes to student debt." Review of. *Journal of economic psychology* 16 (4):663-79.

De Almeida Filho, Adiel T, Christophe Mues, and Lyn C Thomas. 2010. "Optimizing the collections process in consumer credit." Review of. *Production and Operations Management* 19 (6):698-708.

De Servigny, Arnaud, and Olivier Renault. 2002. "Default correlation: empirical evidence." Review of. *Standard and Poor's*.

Deng, Yongheng, John M Quigley, Robert Van Order, and Freddie Mac. 1996. "Mortgage default and low downpayment loans: the costs of public subsidy." Review of. *Regional science and urban economics* 26 (3):263-85.

Dermine, Jean, and C Neto De Carvalho. 2006. "Bank loan losses-given-default: A case study." Review of. *Journal of Banking & Finance* 30 (4):1219-43.

———. 2008. "Bank loan-loss provisioning, central bank rules vs. estimation: The case of Portugal." Review of. *Journal of Financial Stability* 4 (1):1-22.

Dey, Shubhasis, Ramdane Djoudad, and Yaz Terajima. 2008. "Household Indebtedness and Delinquency: Simulations and Estimations Using Microdata." Review of.

Diacon, Stephen. 2004. "Investment risk perceptions: Do consumers and advisers agree?" Review of. *International Journal of Bank Marketing* 22 (3):180-99.

Diacon, Stephen, and Christine Ennew. 2001. "Consumer perceptions of financial risk." Review of. *Geneva Papers on Risk and Insurance. Issues and Practice*:389-409.

Dickey, David A, and Wayne A Fuller. 1979. "Distribution of the estimators for autoregressive time series with a unit root." Review of. *Journal of the American statistical association* 74 (366a):427-31.

Ding, Cherng G, Kuochung Chang, and Na-Ting Liu. 2009. "The roles of personality and general ethical judgments in intention to not repay credit card expenses." Review of. *The Service Industries Journal* 29 (6):813-34.

Duffie, Darrell. 1998. "Defaultable term structure models with fractional recovery of par." In.: Working Paper, Graduate School of Business, Stanford University.

Einhorn, Hillel J, and Robin M Hogarth. 1981. "Behavioral decision theory: Processes of judgment and choice." Review of. *Journal of Accounting Research*:1-31.

Eom, Young Ho, Jean Helwege, and Jing-zhi Huang. 2004. "Structural models of corporate bond pricing: An empirical analysis." Review of. *The Review of Financial Studies* 17 (2):499-544.

Ferrari, Silvia, and Francisco Cribari-Neto. 2004. "Beta regression for modelling rates and proportions." Review of. *Journal of Applied Statistics* 31 (7):799-815.

Finlay, Steven. 2008. *The management of consumer credit: theory and practice*: Palgrave.

———. 2010. "Credit scoring for profitability objectives." Review of. *European Journal of Operational Research* 202 (2):528-37.

Franks, Julian R, and Walter N Torous. 1994. "A comparison of financial recontracting in distressed exchanges and Chapter 11 reorganizations." Review of. *Journal of financial economics* 35 (3):349-70.

Friedman, Craig, and Sven Sandow. 2003. "Recovery rates: Ultimate recoveries." Review of. *RISK-LONDON-RISK MAGAZINE LIMITED-* 16 (8):69-73.

Friedman, Milton. 1957. "The permanent income hypothesis." In *A theory of the consumption function*, 20-37. Princeton University Press.

Frontczak, Robert, and Stefan Rostek. 2015. "Modeling loss given default with stochastic collateral." Review of. *Economic Modelling* 44:162-70.

Geske, Robert. 1977. "The valuation of corporate liabilities as compound options." Review of. *Journal of Financial and quantitative Analysis* 12 (4):541-52.

Gordy, Michael B. 2000. "A comparative anatomy of credit risk models." Review of. *Journal of Banking & Finance* 24 (1):119-49.

Gross, David B, and Nicholas S Souleles. 2002. "Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data." Review of. *The Quarterly journal of economics* 117 (1):149-85.

Grunert, Jens, and Martin Weber. 2005. "Recovery rates of bank loans: Empirical evidence for Germany." Review of.

Gujarati, Damodar N, and Dawn C Porter. 2003. "Basic Econometrics. 4th." In.: New York: McGraw-Hill.

Gupton, Greg M. 2005. "Advancing loss given default prediction models: how the quiet have quickened." Review of. *Economic Notes* 34 (2):185-230.

Gupton, Greg M, Roger M Stein, André Salaam, and David Bren. 2002. "LOSSCALCTM: Model for predicting loss given default (LGD)." Review of. *Moody's KMV, New York*.

Gürtler, Marc, and Martin Hibbeln. 2013. "Improvements in loss given default forecasts for bank loans." Review of. *Journal of Banking & Finance* 37 (7):2354-66.

Harrell, Frank E. 2001. "Regression modeling strategies, with applications to linear models, survival analysis and logistic regression." Review of. *GET ADDRESS: Springer*.

Hoechstoetter, Markus, Abdolreza Nazemi, Svetlozar T Rachev, and Caslav Bozic. 2012. "Recovery Rate Modelling of Non-performing Consumer Credit Using Data Mining Algorithms." Review of.

Hull, John, and Alan White. 1995. "The impact of default risk on the prices of options and other derivative securities." Review of. *Journal of Banking & Finance* 19 (2):299-322.

Huston, Sandra J. 2010. "Measuring financial literacy." Review of. *Journal of Consumer Affairs* 44 (2):296-316.

II, Basel. 2004. "International Convergence of Capital Measurement and Capital Standards: a Revised Framework Basle Committee on Banking Supervision." Review of.

Jacobs Jr, Michael, and Ahmet K Karagozoglu. 2011. "Modeling ultimate loss given default on corporate debt." Review of. *The Journal of Fixed Income* 21 (1):6-20.

Jarrow, Robert A, David Lando, and Stuart M Turnbull. 1997. "A Markov model for the term structure of credit risk spreads." Review of. *The Review of Financial Studies* 10 (2):481-523.

Jarrow, Robert A, and Stuart M Turnbull. 1995. "Pricing derivatives on financial securities subject to credit risk." Review of. *The Journal of finance* 50 (1):53-85.

Ji, Tingting. 2004. "Consumer Credit Delinquency and Bankruptcy Forecasting Using Advanced Econometrc Modeling." Review of.

Jokivuolle, Esa, and Matti Viren. 2013. "Cyclical default and recovery in stress testing loan losses." Review of. *Journal of Financial Stability* 9 (1):139-49.

Katona, George. 1975. *Psychological economics*: Elsevier.

Kau, James B, and Donald C Keenan. 1995. "An overview of the option-theoretic pricing of mortgages." Review of. *Journal of Housing Research*:217-44.

Kavvathas, Dimitrios. 2001. "Estimating credit rating transition probabilities for corporate bonds." Review of.

Kennedy, Peter. 2008. "A guide to modern econometrics." In.: Oxford: Blackwell Publishing.

Khieu, Hinh D, Donald J Mullineaux, and Ha-Chin Yi. 2012. "The determinants of bank loan recovery rates." Review of. *Journal of Banking & Finance* 36 (4):923-33.

Kim, In Joon, Krishna Ramaswamy, and Suresh Sundaresan. 1993. "Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model." Review of. *Financial Management*:117-31.

Koopman, Siem Jan, and Andre Lucas. 2005. "Business and default cycles for credit risk." Review of. *Journal of Applied Econometrics* 20 (2):311-23.

Koopman, Siem Jan, André Lucas, and André Monteiro. 2008. "The multi-state latent factor intensity model for credit rating transitions." Review of. *Journal of Econometrics* 142 (1):399-424.

Lambrecht, Bart, William Perraudin, and Stephen Satchell. 1997. "Time to default in the UK mortgage market." Review of. *Economic Modelling* 14 (4):485-99.

Lando, David. 1998. "On Cox processes and credit risky securities." Review of. *Review of Derivatives research* 2 (2-3):99-120.

Lea, Stephen EG, Paul Webley, and Catherine M Walker. 1995. "Psychological factors in consumer debt: Money management, economic socialization, and credit use." Review of. *Journal of economic psychology* 16 (4):681-701.

Leland, Hayne E, and Klaus Bjerre Toft. 1996. "Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads." Review of. *The Journal of finance* 51 (3):987-1019.

Litterman, Robert B, and Thomas Iben. 1991. "Corporate bond valuation and the term structure of credit spreads." Review of. *The journal of portfolio management* 17 (3):52-64.

Livingstone, Sonia M, and Peter K Lunt. 1992. "Predicting personal debt and debt repayment: Psychological, social and economic determinants." Review of. *Journal of economic psychology* 13 (1):111-34.

Longstaff, Francis A, and Eduardo S Schwartz. 1995. "A simple approach to valuing risky fixed and floating rate debt." Review of. *The Journal of finance* 50 (3):789-819.

Lucas, Alan. 2006. Basel II problem solving. Paper presented at the Conference on Basel II and credit risk modelling in consumer lending.

Lusardi, Annamaria, and Olivia S Mitchell. 2008. "Planning and financial literacy: How do women fare?" In.: National Bureau of Economic Research.

———. 2014. "The economic importance of financial literacy: Theory and evidence." Review of. *Journal of Economic Literature* 52 (1):5-44.

Lusardi, Annamaria, and Peter Tufano. 2009. "Debt literacy, financial experiences, and overindebtedness." In.: National Bureau of Economic Research.

Madan, Dileep, and Haluk Unal. 1995. "Pricing the risk of recovery in default with apr violation." Review of. *Journal of Banking and Finance* 27 (6):1001-218.

Matuszyk, Ania, Christophe Mues, and Lyn C Thomas. 2010. "Modelling LGD for unsecured personal loans: Decision tree approach." Review of. *Journal of the Operational Research Society* 61 (3):393-8.

Memmel, Christoph, Yalin Gündüz, and Peter Raupach. 2015. "The common drivers of default risk." Review of. *Journal of Financial Stability* 16:232-47.

Merton, Robert C. 1974. "On the pricing of corporate debt: The risk structure of interest rates." Review of. *The Journal of finance* 29 (2):449-70.

Nagpal, Krishan, and Reza Bahar. 2001. "Measuring default correlation." Review of. *Risk* 14 (3):129-32.

Newey, Whitney K, and Kenneth D West. 1987. "Hypothesis testing with efficient method of moments estimation." Review of. *International Economic Review*:777-87.

Nickell, Pamela, William Perraudin, and Simone Varotto. 2000. "Stability of rating transitions." Review of. *Journal of Banking & Finance* 24 (1):203-27.

Norvilitis, Jill M, Michelle M Merwin, Timothy M Osberg, Patricia V Roehling, Paul Young, and Michele M Kamas. 2006a. "Personality factors, money attitudes, financial knowledge, and credit ‐ card debt in college students." Review of. *Journal of Applied Social Psychology* 36 (6):1395-413.

———. 2006b. "Personality factors, money attitudes, financial knowledge, and credit ‐ card debt in college students1." Review of. *Journal of Applied Social Psychology* 36 (6):1395-413.

Norvilitis, Jill M, P Bernard Szablicki, and Sandy D Wilson. 2003. "Factors Influencing Levels of Credit ‐ Card Debt in College Students1." Review of. *Journal of Applied Social Psychology* 33 (5):935-47.

O'sullivan, AJ, TR Cheek, RB Moreton, MJ Berridge, and RD Burgoyne. 1989. "Localization and heterogeneity of agonist-induced changes in cytosolic calcium concentration in single bovine adrenal chromaffin cells from video imaging of fura-2." Review of. *The EMBO Journal* 8 (2):401.

Ospina, R, and SLP Ferrati. 2010. "Inflated beta regression models." In.: Working paper, Universidade Federal de Pernambuco and Universidade de Sao Paulo.

Ospina, Raydonal, and Silvia LP Ferrari. 2012. "A general class of zero-or-one inflated beta regression models." Review of. *Computational Statistics & Data Analysis* 56 (6):1609-23.

Pacter, Paul. 2014. *IFRS as global standards: a pocket guide*: IFRS Foundation.

Patterson, Kerry D. 2000. *An introduction to applied econometrics*: St. Martin's press.

Perry, Vanessa Gail. 2008. "Giving credit where credit is due: The psychology of credit ratings." Review of. *The Journal of Behavioral Finance* 9 (1):15-21.

Phillips, Peter CB, and Pierre Perron. 1988. "Testing for a unit root in time series regression." Review of. *Biometrika* 75 (2):335-46.

Ramirez, Juan. 2015. *Accounting for derivatives: Advanced hedging under IFRS 9*: John Wiley & Sons.

Renault, Olivier, and Olivier Scaillet. 2004. "On the way to recovery: A nonparametric bias free estimation of recovery rate densities." Review of. *Journal of Banking & Finance* 28 (12):2915-31.

Repullo, Rafael, Jesús Saurina, and Carlos Trucharte. 2010. "Mitigating the Pro-cyclicality of Basel II." Review of. *Economic Policy* 25 (64):659-702.

Repullo, Rafael, and Javier Suarez. 2008. "The procyclical effects of Basel II." Review of. *Vol*.

Rottenstreich, Yuval, and Amos Tversky. 1997. "Unpacking, repacking, and anchoring: advances in support theory." Review of. *Psychological review* 104 (2):406.

Siddiqi, Naeem. 2006. "Credit risk scorecards." Review of. *Developing and Implementing Intelligent Credit Scoring*.

Singleton, Vernon L, Rudolf Orthofer, and Rosa M Lamuela-Raventós. 1999. "[14] Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent." Review of. *Methods in enzymology* 299:152-78.

Stiglitz, Joseph E, and Andrew Weiss. 1981. "Credit rationing in markets with imperfect information." Review of. *The American economic review* 71 (3):393-410.

Thomas, Hugh, and Zhiqiang Wang. 2005. "Interpreting the internal ratings-based capital requirements in Basel II." Review of. *Journal of Banking Regulation* 6 (3):274-89.

Thomas, LC, A Matuszyk, and A Moore. 2012. "Comparing debt characteristics and LGD models for different collections policies." Review of. *International Journal of Forecasting* 28 (1):196-203.

Thomas, Lyn C, Anna Matuszyk, Mee Chi So, Christophe Mues, and Angela Moore. 2016. "Modelling repayment patterns in the collections process for unsecured consumer debt: A case study." Review of. *European Journal of Operational Research* 249 (2):476-86.

Tokunaga, Howard. 1993. "The use and abuse of consumer credit: Application of psychological theory and research." Review of. *Journal of economic psychology* 14 (2):285-316.

Trück, Stefan, and Svetlozar Rachev. 2005. "Credit portfolio risk and PD confidence sets through the business cycle." Review of. *Available at SSRN 675622*.

Vasicek, Oldrich A. 1987. *Probability of loss on loan portfolio*: Wiley Online Library.

Vio, BB. 2008. "Credit beyond rationality. 2008. 56p." Dissertation of Master of Science–Advanced Study in Economic and Consumer Psychology of University of Exeter, Exeter.[Links].

Wallach, Michael A, and Nathan Kogan. 1965. "The roles of information, discussion, and consensus in group risk taking." Review of. *Journal of Experimental Social Psychology* 1 (1):1-19.

Yao, Xiao, Jonathan Crook, and Galina Andreeva. 2015. "Support vector regression for loss given default modelling." Review of. *European Journal of Operational Research* 240 (2):528-38.

Zakrajsek, Egon, Seth B Carpenter, and William C Whitesell. 2001. "Capital requirements, business loans, and business cycles: an empirical analysis of the standardized approach in the New Basel Capital Accord." Review of.

Zhang, Jie, and Lyn C Thomas. 2009. "Comparison of single distribution and mixture distribution models for modelling LGD." Review of.

———. 2012. "Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD." Review of. *International Journal of Forecasting* 28 (1):204-15.

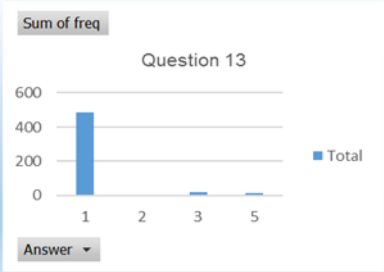
Appendices

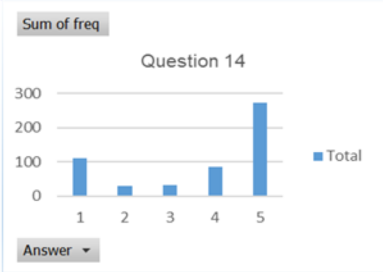
Appendix 1. Questionnaire Questions and Expected Answers

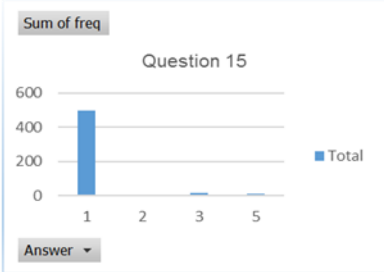
Number	Question	Expected Value
	Financial Knowledge	
1	The lender has the right to send the borrower's name to a negative credit bureau.	1
2	When in default, borrowers should not be permitted to take out further credit s specified period of time.	1
2.1	If true, how long?	5
3	If the chance of getting a disease is 10%, how many people out of 100 would be expected to get the disease?	10
4	If 5 people each have the winning number in the lottery and the prize is £ 2,000.00, how much will each of them receive?	400
5	If you have £ 200.00 in savings account, which earns 2% interest rates per month, how much will be in the account at the end of one year?	255
6	If you take a loan of £ 500.00 and the lender asks you to pay £ 75.00 in advance, how much will the outstanding amount be?	425
7	We are interested in what is most important for you when you apply for a loan. Please rate the following statement with a score from 1 to 3, in which 1 represents the most import and 3 the least important. (1) Instalment's Value (2) Interest Rate (3) Loan's Length	123 132 213 231 312 321
	Personal Finance behaviour	
8	In the past three years how many loans have you had taken?	1 (yes) 2 (no)
9	Thinking of the last 6 month, have you faced one of the possible situation below? a. Health problem b. Unemployment c. Unexpected bills d. Divorce e. Children birth f. Unexpected car expenses g. Unexpected expenses with another family member h. None	 yes/no yes/no yes/no yes/no yes/no yes/no yes/no

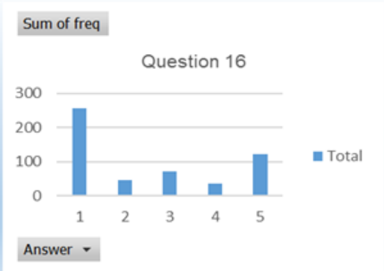
		yes/no
10	Do you have private health insurance?	yes/no
11	Have you ever taken credit for another person?	yes/no
Risk Taking rated with a score from 1 to 5, where 1 represents disapprove and 5 approve		
12	Quitting a job that they dislike without having a new one lined up	1-5
13	Swimming far out from the seashore	1-5
14	Investing 10% of their annual income in a moderate growth mutual fund	1-5
15	Drinking heavily at a social function	1-5
16	Approaching their boss for a rise	1-5
17	Wearing provocative or unconventional clothes	1-5
18	Riding a motorcycle without a helmet	1-5
19	Investing 10% of their annual income in a new business venture	1-5
20	Downloading property software from the internet	1-5
21	Keeping a wallet that contains £ 200.00	1-5
Demographic Information		
22	Education level a. Primary school b. High school c. Graduate d. Post Graduate	yes/no yes/no yes/no yes/no
23	Residential Status a. Owner with mortgage b. Owner c. Rent d. Others	yes/no yes/no yes/no yes/no
24	Religion a. Catholic b. Protestant c. Spiritualism d. African e. Others	yes/no yes/no yes/no yes/no yes/no
25	Marital Status a. Married b. Single c. Divorced d. Others	yes/no yes/no yes/no yes/no

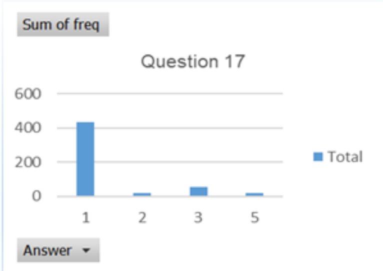
Appendix 2. Questionnaire Answer Histograms

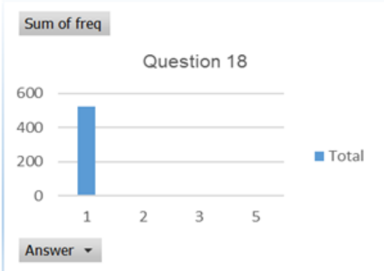

Sum of freq
Question 13
600
400
200
0
1 2 3 5
Total
Answer

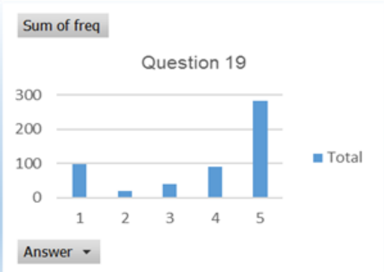

Sum of freq
Question 14
300
200
100
0
1 2 3 4 5
Total
Answer


Sum of freq
Question 15
600
400
200
0
1 2 3 5
Total
Answer


Sum of freq
Question 16
300
200
100
0
1 2 3 4 5
Total
Answer

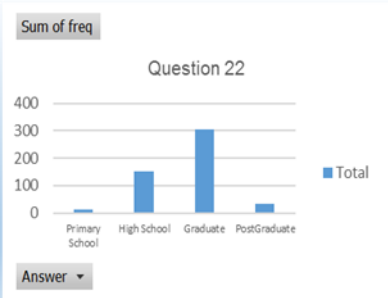

Sum of freq
Question 17
600
400
200
0
1 2 3 5
Total
Answer

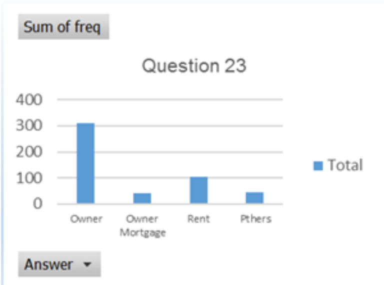

Sum of freq
Question 18
600
400
200
0
1 2 3 5
Total
Answer


Sum of freq
Question 19
300
200
100
0
1 2 3 4 5
Total
Answer

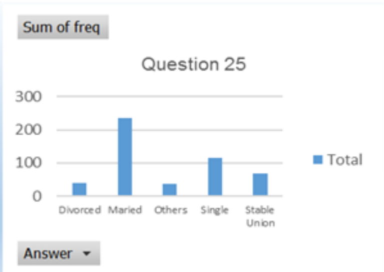

Sum of freq
Question 20
600
400
200
0
1 2 3 4 5
Total
Answer


Sum of freq
Question 21
600
400
200
0
1 2 3 4 5
Total
Answer


Sum of freq
Question 22
400
300
200
100
0
Primary School
High School
Graduate
PostGraduate
Total
Answer


Sum of freq
Question 23
400
300
200
100
0
Owner
Owner Mortgage
Rent
Pthers
Total
Answer


Sum of freq
Question 24
300
200
100
0
African
Catholic
Others
Protestant
Spiritualist
Total
Answer


Sum of freq
Question 25
300
200
100
0
Divorced
Maried
Others
Single
Stable Union
Total
Answer

Appendix 3. Original Variables Names and Frequencies

Category	Variable Name	Description	Frequency
Borrowers' gender	d_gen	Female	201
	ref*	Male	263
Borrowers' age	d_age_1	<=30	41
	d_age_2	31-40	120
	d_age_3	41-50	136
	d_age_4	51-60	71
	d_age_5	61-70	56
	ref*	>70	40
Borrowers' income	d_sal_1	<=600	24
	d_sal_2	>600 - <=1200	207
	d_sal_3	>1200 - <=2000	163
	d_sal_4	>2000 - <=2800	33
	ref*	>2800	37
Borrowers' employment length	d_emp_time_1	<12	151
	d_emp_time_2	>=12 - <14	108
	d_emp_time_3	>=14 - <20	115
	d_emp_time_4	>=20 - <25	46
	d_emp_time_5	>=25 - <30	20
	ref*	>= 30	24
Products	d_prod_1	Sound & video	39
	d_prod_2	Appliance	77
	d_prod_3	Mobile phone	145
	d_prod_4	Small appliance	36
	d_prod_5	Furniture	127
	d_prod_6	Computing	26
	ref*	Sport & Leisure	14
Monthly Loan Interest Rate	d_int_1	4.9%	57
	d_int_2	5.0%	16
	d_int_3	5.4%	37
	d_int_4	5.8%	217
	d_int_5	6.1%	69
	d_int_6	6.8%	26
	ref*	7.0%	42
Ratio income committed / instalment	d_dti_1	<=0.5	73
	d_dti_2	>0.5 - <=1.0	121
	d_dti_3	>1.0 - <=1.5	90

Category	Variable Name	Description	Frequency
	d_dti_4	>1.5 - <=2.0	57
	d_dti_5	>2.0 - <=2.5	50
	ref*	>2.5	73
Ratio income committed / loan amount	d_lti_1	<=0.05	55
	d_lti_2	>0.05 - <=0.10	140
	d_lti_3	>0.10 - <=0.15	137
	d_lti_4	>0.15 - <=0.20	68
	d_lti_5	>0.20 - <=0.25	31
	ref*	>0.25	33
Borrowers' marital status	d_mst_1	Single	102
	d_mst_2	Married	197
	d_mst_3	Stable union	63
	d_mst_4	Divorced	36
	ref*	Others	66
Borrowers' religion	d_rel_1	Catholic	236
	d_rel_2	Protestant	133
	d_rel_3	Spiritualist	13
	d_rel_4	African	8
	ref*	Others	72
Borrowers' home status	d_hst_1	Owner	262
	d_hst_2	Owner mortgage	40
	d_hst_3	Rent	93
	ref*	Others	69
Borrowers' education level	d_edu_1	Primary School	11
	d_edu_2	High School	127
	d_edu_3	Graduate	268
	ref*	Post Graduate	58
Reasons for Missing Payment	d_rea_1	Health problem	270
	d_rea_2	Unemployment	99
	d_rea_3	Unexpected bills	48
	d_rea_4	Divorce	8
	d_rea_5	Children birth	13
	d_rea_6	Car expenses	18
	ref*	Expenses family	8
Financial Knowledge Question 1	d_cor_1	Correct	387
	ref*	Incorrect	77
Financial Knowledge Question 2	d_cor_2	Correct	409
	ref*	Incorrect	55

Category	Variable Name	Description	Frequency
Financial Knowledge Question 21	d_cor_21	Correct	149
	ref*	Incorrect	315
Financial Knowledge Question 3	d_cor_3	Correct	230
	ref*	Incorrect	234
Financial Knowledge Question 4	d_cor_4	Correct	131
	ref*	Incorrect	333
Financial Knowledge Question 5	d_cor_5	Correct	2
	ref*	Incorrect	462
Financial Knowledge Question 6	d_cor_6	Correct	261
	ref*	Incorrect	203
Financial Knowledge Question 7 1-Instalment amount 2-Loan interest rate 3-Contract length	d_cor7_1	123	83
	d_cor7_2	132	23
	d_cor7_3	213	146
	d_rea7_4	231	134
	d_cor7_5	312	37
	ref*	321	41
Financial Knowledge Degree of accuracy	qdist01	[0,1]	464
	qdist02	[0,1]	464
	qdist021	[0,5]	464
	qdist03	[0,990]	464
	qdist04	[0.9600]	464
	qdist05	[0,255]	464
	qdist06	[0,425]	464
Financial Knowledge	Score	Continuous (0,1)	464
Risk Taking	d_12_1	Strong disapprove	381
	d_12_2	Disapprove	16
	d_12_3	Neutral	20
	d_12_4	Approve	11
	ref*	Strong approve	36
	d_13_1	Strong disapprove	428
	d_13_2	Disapprove	6
	d_13_3	Neutral	17
	d_13_4	Approve	0
	ref*	Strong approve	13
	d_14_1	Strong disapprove	100
	d_14_2	Disapprove	26
	d_14_3	Neutral	26
	d_14_4	Approve	80

Category	Variable Name	Description	Frequency
	ref*	Strong approve	232
	d_15_1	Strong disapprove	439
	d_15_2	Disapprove	3
	d_15_3	Neutral	15
	d_15_4	Approve	0
	ref*	Strong approve	7
	d_16_1	Strong disapprove	231
	d_16_2	Disapprove	44
	d_16_3	Neutral	55
	d_16_4	Approve	34
	ref*	Strong approve	100
	d_17_1	Strong disapprove	390
	d_17_2	Disapprove	16
	d_17_3	Neutral	42
	d_17_4	Approve	0
	ref*	Strong approve	16
	d_18_1	Strong disapprove	458
	d_18_2	Disapprove	3
	d_18_3	Neutral	2
	d_18_4	Approve	0
	ref*	Strong approve	1
	d_19_1	Strong disapprove	87
	d_19_2	Disapprove	20
	d_19_3	Neutral	33
	d_19_4	Approve	84
	ref*	Strong approve	240
	d_20_1	Strong disapprove	366
	d_20_2	Disapprove	8
	d_20_3	Neutral	33
	d_20_4	Approve	3
	ref*	Strong approve	54
	d_21_1	Strong disapprove	366
	d_21_2	Disapprove	8
	d_21_3	Neutral	33
	d_21_4	Approve	3
	ref*	Strong approve	54

Appendix 4. Transformed Variables Names and Frequencies

Category	Variable Name	Description	Frequency
Borrowers' gender	d_gen	Female	201
	ref*	Male	263
Borrowers' age	d_age_1		
	d_age_2		
	d_age_3		1
	ref*		
Borrowers' income	d_sal_1	<=600	24
	d_sal_2	>600 - <=1200	207
	d_sal_3	>1200 - <=2000	163
	ref*	>2800	37
Borrowers' employment length	d_emp_time_1	<12	151
	d_emp_time_2	>=12 - <14	108
	d_emp_time_3	>=14 - <20	115
	ref*	>= 30	24
Products	d_prod_1	Sound & video	39
	d_prod_2	Appliance	77
	d_prod_3	Mobile phone	145
	ref*	Sport & Leisure	14
Monthly Loan Interest Rate	d_inter_1	4.9%	57
	d_inter_2	5.0%	16
	d_inter_3	5.4%	37
	ref*	7.0%	42
Ratio income committed / instalment	d_dti_1	<=0.5	73
	d_dti_2	>0.5 - <=1.0	121
	d_dti_3	>1.0 - <=1.5	90
	d_dti_4	>1.5 - <=2.0	57
	ref*	>2.5	73
Ratio income committed / loan amount	d_lti_1	<=0.05	55
	d_lti_2	>0.05 - <=0.10	140
	d_lti_3	>0.10 - <=0.15	137
	d_lti_4	>0.15 - <=0.20	68
	ref*	>0.25	33
Borrowers' marital status	d_mst_1	Single	102
	ref*	Others	66
Borrowers' religion	d_rel_1	Catholic	236
	d_rel_2	Protestant	133

	ref*	Others	72
Borrowers' home status	d_hst_1	Owner	262
	ref*	Others	69
Borrowers' education level	d_edu_1	Primary School	11
	ref*	Post Graduate	58
Reasons for Missing Payment	d_rea_1	Health problem	270
	d_rea_2	Unemployment	99
	ref*	Expenses family	8
Financial Knowledge Question 1	d_cor_1	Correct	387
	ref*	Incorrect	77
Financial Knowledge Question 2	d_cor_2	Correct	409
	ref*	Incorrect	55
Financial Knowledge Question 21	d_cor_21	Correct	149
	ref*	Incorrect	315
Financial Knowledge Question 3	d_cor_3	Correct	230
	ref*	Incorrect	234
Financial Knowledge Question 4	d_cor_4	Correct	131
	ref*	Incorrect	333
Financial Knowledge Question 5	d_cor_5	Correct	2
	ref*	Incorrect	462
Financial Knowledge Question 6	d_cor_6	Correct	261
	ref*	Incorrect	203
Financial Knowledge Question 7 1-Instalment amount 2-Loan interest rate 3-Contract length	d_cor7_1	123	83
	d_cor7_2	132	23
	d_cor7_3	213	146
	d_rea7_4	231	134
	d_cor7_5	312	37
	ref*	321	41
Financial Knowledge Degree of accuracy	cdist01		464
	cdist02		464
	cdist021		464
	cdist03		464
	cdist04		464
	cdist05		464
	cdist06		464
Financial Knowledge	Score	Continuous (0,1)	464
Risk Taking	d_12_1	Disapprove	
	d_12_2	Approve	
	ref*	Neutral	

	d_13_1	Disapprove	
	d_13_2	Approve	
	ref*	Neutral	
	d_14_1	Disapprove	
	d_14_2	Approve	
	ref*	Neutral	
	d_15_1	Disapprove	
	d_15_2	Approve	
	ref*	Neutral	
	d_16_1	Disapprove	
	d_16_2	Approve	
	ref*	Neutral	
	d_17_1	Disapprove	
	d_17_2	Approve	
	ref*	Neutral	
	d_18_1	Disapprove	
	d_18_2	Approve	
	ref*	Neutral	
	d_19_1	Disapprove	
	d_19_2	Approve	
	ref*	Neutral	
	d_20_1	Disapprove	
	d_20_2	Approve	
	ref*	Neutral	
	d_21_1	Disapprove	
	d_21_2	Approve	
	ref*	Neutral	

Appendix 5. Model OLS O1 and OLS O2

Dependent Variable: Recovery Rate				
	OLS O1		OLS O2	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	0.7961**	0.0251	0.8516**	0.0195
d_Gen	0.0102	0.7309	-0.00444	0.885
d_Age_1	-0.1181	0.1167	-0.0712	0.3877
d_Age_2	-0.1048*	0.0932	-0.0676	0.3351
d_Age_3	-0.0925	0.1168	-0.0607	0.35
d_Age_4	-0.0827	0.1853	-0.0542	0.4236
d_Age_5	-0.0266	0.6722	-0.0131	0.8427
Variables	Estimate	P-Value	Estimate	P-Value
d_emp_time_1	-0.0573	0.4055	-0.0455	0.5142
d_emp_time_2	-0.0705	0.3082	-0.0663	0.3416
d_emp_time_3	-0.0627	0.3525	-0.0453	0.5073
d_emp_time_4	-0.1429*	0.0596	-0.1341*	0.0847
d_emp_time_5	-0.0917	0.305	-0.07	0.4424
d_sal_1	0.1274	0.161	0.1099	0.2353
d_sal_2	0.15**	0.0174	0.1451**	0.0228
d_sal_3	0.0699	0.2411	0.0678	0.2597
d_sal_4	0.0992	0.1681	0.0912	0.2145
d_prod_1	-0.0877	0.7731	-0.1738	0.5751
d_prod_2	-0.0113	0.9702	-0.0887	0.7724
d_prod_3	-0.0953	0.7508	-0.1803	0.5551
d_prod_4	-0.1483	0.6365	-0.2354	0.461
d_prod_5	-0.0993	0.7414	-0.1737	0.57
d_prod_6	-0.0259	0.9318	-0.1094	0.722
d_prod_7	-0.1491	0.6318	-0.2512	0.4289
d_dti_1	0.00912	0.9259	-0.0307	0.7635
d_dti_2	-0.0504	0.5387	-0.0861	0.3106
d_dti_3	-0.0463	0.5321	-0.0804	0.2926
d_dti_4	-0.00391	0.9562	-0.0297	0.6851
d_dti_5	-0.0805	0.2804	-0.1059	0.1618
d_lti_1	0.1316*	0.079	0.1464*	0.0581
d_lti_2	0.1702***	0.0086	0.187***	0.0051
d_lti_3	0.1597***	0.0053	0.161***	0.0061
d_lti_4	0.1449**	0.0187	0.1571**	0.013
d_lti_5	0.1125*	0.0588	0.112*	0.0693

Dependent Variable: Recovery Rate				
	OLS O1		OLS O2	
d_int_1	-0.1347**	0.0364	-0.1229*	0.0599
d_int_2	-0.5295***	<.0001	-0.5289***	<.0001
d_int_3	0.0294	0.6586	0.0239	0.7243
d_int_4	-0.1163**	0.033	-0.1176**	0.0325
d_int_5	-0.427***	<.0001	-0.4326***	<.0001
d_int_6	-0.2399***	0.0018	-0.2429***	0.0017
d_mst_1			0.00124	0.9816
d_mst_2			0.00316	0.9488
d_mst_3			0.0103	0.8636
d_mst_4			0.0724	0.3039
Variables	Estimate	P-Value	Estimate	P-Value
d_rel_1			0.0551	0.2534
d_rel_2			0.0707	0.1654
d_rel_3			0.1206	0.1851
d_rel_4			0.2063*	0.0741
d_hst_1			-0.017	0.7164
d_hst_2			-0.1014	0.1124
d_hst_3			-0.0651	0.22
d_edu_0			0.0623	0.5698
d_edu_1			-0.0436	0.4353
d_edu_2			-0.0181	0.7253

Appendix 6. Model OLS O3 and OLS O3

Dependent Variable: Recovery Rate				
	OLS O3		OLS O4	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	1.0319***	0.0061	1.0002**	0.0244
d_Gen	-0.00309	0.9195	-0.00267	0.9333
d_Age_1	-0.0587	0.4802	-0.0682	0.4263
d_Age_2	-0.0657	0.3511	-0.0697	0.3355
d_Age_3	-0.0646	0.3242	-0.0642	0.3379
d_Age_4	-0.0475	0.4804	-0.0515	0.4552
d_Age_5	-0.0134	0.8375	-0.0139	0.836
d_emp_time_1	-0.0251	0.7207	-0.0259	0.7192
d_emp_time_2	-0.0473	0.5008	-0.0473	0.5122
d_emp_time_3	-0.0365	0.5924	-0.0379	0.5861
d_emp_time_4	-0.1208	0.1191	-0.1328	0.0932
d_emp_time_5	-0.055	0.541	-0.0506	0.5832
d_sal_1	0.1381	0.134	0.144	0.1315
d_sal_2	0.1602**	0.0113	0.1612**	0.0126
d_sal_3	0.0804	0.1798	0.0844	0.1718
d_sal_4	0.092	0.2035	0.0942	0.2058
d_prod_1	-0.2721	0.3748	-0.2643	0.3977
d_prod_2	-0.1893	0.5327	-0.1787	0.5642
d_prod_3	-0.2911	0.3353	-0.2884	0.3494
d_prod_4	-0.3643	0.2501	-0.3561	0.269
d_prod_5	-0.2874	0.342	-0.2847	0.3557
d_prod_6	-0.218	0.4726	-0.2173	0.483
d_prod_7	-0.358	0.254	-0.3441	0.2815
d_dti_1	-0.0144	0.8878	-0.0178	0.8665
d_dti_2	-0.0865	0.3069	-0.0886	0.3073
d_dti_3	-0.0892	0.2408	-0.097	0.2152
d_dti_4	-0.0366	0.6177	-0.04	0.599
d_dti_5	-0.1162	0.1219	-0.1192	0.1214
d_lti_1	0.161**	0.0355	0.1646**	0.0357
d_lti_2	0.1941***	0.0033	0.199***	0.0034
d_lti_3	0.1767***	0.0023	0.1787***	0.0027
d_lti_4	0.1829***	0.0034	0.1976***	0.0021
d_lti_5	0.1192*	0.0513	0.1176*	0.0621
d_int_1	-0.1344**	0.0387	-0.1336**	0.0469
d_int_2	-0.546***	<.0001	-0.5418***	<.0001
d_int_3	0.00509	0.9404	0.00429	0.951
d_int_4	-0.1274**	0.0197	-0.1251**	0.0268
d_int_5	-0.4393***	<.0001	-0.4439***	<.0001
d_int_6	-0.249***	0.0012	-0.246***	0.0018
d_mst_1	0.0181	0.7372	0.0139	0.8012

Dependent Variable: Recovery Rate				
	OLS O3		OLS O4	
Variables	Estimate	P-Value	Estimate	P-Value
d_mst_2	0.00449	0.927	0.00223	0.9648
d_mst_3	0.0212	0.721	0.0148	0.8077
d_mst_4	0.0686	0.3271	0.0669	0.3494
d_rel_1	0.0656	0.1699	0.0559	0.2559
d_rel_2	0.0789	0.1161	0.0673	0.1933
d_rel_3	0.137	0.1267	0.1418	0.1263
d_rel_4	0.2326**	0.0413	0.2337**	0.0439
d_hst_1	-0.0153	0.7394	-0.0199	0.6725
d_hst_2	-0.1074*	0.0875	-0.1105*	0.0869
d_hst_3	-0.0492	0.3502	-0.044	0.4101
d_edu_0	0.0594	0.5819	0.074	0.5072
d_edu_1	-0.0504	0.3619	-0.042	0.4743
d_edu_2	-0.0299	0.558	-0.0272	0.606
d_rea_1	-0.1133	0.3172	-0.1079	0.3519
d_rea_2	-0.1843	0.1174	-0.1793	0.1367
d_rea_3	-0.0629	0.5953	-0.05	0.6792
d_rea_4	-0.165	0.2699	-0.1614	0.2907
d_rea_5	-0.0931	0.4988	-0.0759	0.5938
d_rea_6	0.1285	0.3354	0.1304	0.3377
d_cor1			0.0251	0.5344
d_cor2			0.0136	0.7735
d_cor21			-0.00446	0.889
d_cor3			-0.0246	0.4543
d_cor4			0.0188	0.5719
d_cor5			0.0322	0.8812
d_cor6			-0.0184	0.5861
d_cor7_1			0.0183	0.7406
d_cor7_2			0.0856	0.2653
d_cor7_3			0.0076	0.8805
d_cor7_4			-0.022	0.6663
d_cor7_5			0.00117	0.9877
***, **, * indicate 1% ,5% and 10% significance levels, respectively.				

Appendix 7. Model OLS O5 and OLS 06

Dependent Variable: Recovery Rate				
	OLS O5		OLS O6	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	1.059***	0.0026	1.0256***	0.0035
d_Gen	0.00462	0.8816	0.00154	0.9605
d_Age_1	-0.0536	0.5061	-0.0605	0.4546
d_Age_2	-0.0457	0.4956	-0.0596	0.3774
d_Age_3	-0.0531	0.392	-0.0571	0.3605
d_Age_4	-0.0333	0.6054	-0.0382	0.5569
d_Age_5	-0.0116	0.8562	-0.00524	0.9351
d_emp_time_1	-0.0312	0.6549	-0.0309	0.6588
d_emp_time_2	-0.0596	0.395	-0.0467	0.5074
d_emp_time_3	-0.0482	0.4785	-0.0425	0.5341
d_emp_time_4	-0.1395*	0.0693	-0.1354*	0.0798
d_emp_time_5	-0.0868	0.3448	-0.064	0.476
d_sal_1	0.1444	0.1165	0.1472	0.1115
d_sal_2	0.1515	0.0172	0.1577**	0.0129
d_sal_3	0.075	0.2158	0.0778	0.1998
d_sal_4	0.0936	0.1941	0.0892	0.2165
d_prod_1	-0.2887	0.3461	-0.2844	0.3547
d_prod_2	-0.2059	0.4979	-0.1993	0.5125
d_prod_3	-0.3159	0.2963	-0.3095	0.3069
d_prod_4	-0.3829	0.2247	-0.3893	0.2182
d_prod_5	-0.3135	0.2985	-0.3136	0.2998
d_prod_6	-0.2465	0.416	-0.2355	0.4385
d_prod_7	-0.3779	0.2287	-0.3805	0.2255
d_dti_1	0.00815	0.9367	-0.00879	0.9317
d_dti_2	-0.0574	0.4955	-0.0775	0.3579
d_dti_3	-0.0733	0.34	-0.088	0.2492
d_dti_4	-0.0141	0.8511	-0.0427	0.5635
d_dti_5	-0.1114	0.1419	-0.111	0.1408
d_lti_1	0.1251	0.1017	0.1587**	0.0381
d_lti_2	0.1744***	0.009	0.2005***	0.0026
d_lti_3	0.1599***	0.006	0.1837***	0.0016
d_lti_4	0.17***	0.007	0.1932***	0.0021
d_lti_5	0.0985	0.1131	0.1257**	0.0407
d_int_1	-0.1439**	0.0288	-0.1265*	0.0535
d_int_2	-0.5575***	<.0001	-0.5429***	<.0001
d_int_3	-0.0157	0.8179	0.00626	0.9269
d_int_4	-0.1301**	0.0174	-0.114**	0.0377
d_int_5	-0.4561***	<.0001	-0.4363***	<.0001
d_int_6	-0.2515***	0.001	-0.2421***	0.0016
d_mst_1	0.0112	0.8343	0.0192	0.7219

Dependent Variable: Recovery Rate				
	OLS O5		OLS O6	
Variables	Estimate	P-Value	Estimate	P-Value
d_mst_2	0.00164	0.9735	0.0042	0.932
d_mst_3	0.0254	0.6685	0.0157	0.7907
d_mst_4	0.073	0.2966	0.0637	0.3648
d_rel_1	0.0563	0.2386	0.0557	0.2451
d_rel_2	0.0725	0.1481	0.073	0.146
d_rel_3	0.1399	0.122	0.1375	0.1291
d_rel_4	0.2281**	0.0432	0.2329**	0.0403
d_hst_1	-0.0191	0.6773	-0.0147	0.749
d_hst_2	-0.0784	0.2132	-0.1006	0.1123
d_hst_3	-0.0452	0.3866	-0.0403	0.4439
d_edu_0	0.0654	0.5483	0.0823	0.4513
d_edu_1	-0.0404	0.474	-0.043	0.4539
d_edu_2	-0.025	0.6258	-0.0268	0.6041
d_rea_1	-0.1351	0.2256	-0.1416	0.207
d_rea_2	-0.2073*	0.0746	-0.2105*	0.0721
d_rea_3	-0.079	0.4985	-0.0822	0.4838
d_rea_4	-0.1868	0.2085	-0.2003	0.1797
d_rea_5	-0.1179	0.3949	-0.1214	0.3807
d_rea_6	0.1022	0.4355	0.1039	0.4316
SQdist01	0.00559	0.6997		
SQdist02	0.00508	0.7378		
SQdist021	0.00306	0.8313		
SQdist03	-0.0653***	0.0015		
SQdist04	0.0106	0.1752		
SQdist05	0.0124	0.4644		
SQdist06	-0.0156	0.3561		
Score			0.0332	0.6254
d_cor7_1	0.0162	0.7641	0.015	0.7816
d_cor7_2	0.0778	0.2961	0.0785	0.2949
d_cor7_3	0.00908	0.853	0.00334	0.9455
d_cor7_4	-0.0123	0.8037	-0.0208	0.6743
d_cor7_5	0.00201	0.9781	-0.00513	0.9439
***, **, * indicate 1% ,5% and 10% significance levels, respectively.				

Appendix 8. Model OLS O7 and OLS O8

Dependent Variable: Recovery Rate				
	OLS O7		OLS O8	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	0.7651	0.1189	0.8082	0.1033
d_Gen	-0.00079	0.9799	0.0029	0.9279
d_Age_1	-0.0921	0.287	-0.0919	0.2901
d_Age_2	-0.0793	0.2588	-0.0639	0.3626
d_Age_3	-0.0732	0.2587	-0.0683	0.2923
d_Age_4	-0.034	0.6119	-0.0278	0.6767
d_Age_5	-0.0325	0.6275	-0.0372	0.5781
d_emp_time_1	-0.045	0.5301	-0.0417	0.5621
d_emp_time_2	-0.0675	0.3492	-0.0789	0.2766
d_emp_time_3	-0.0548	0.4362	-0.0587	0.4057
d_emp_time_4	-0.1514*	0.0561	-0.1601**	0.0434
d_emp_time_5	-0.0749	0.4234	-0.1021	0.2865
d_sal_1	0.1418	0.1391	0.1279	0.1867
d_sal_2	0.1802***	0.0064	0.1655**	0.0133
d_sal_3	0.0944	0.1288	0.0856	0.1732
d_sal_4	0.101	0.1846	0.0973	0.2009
d_prod_1	-0.2067	0.5109	-0.1817	0.5666
d_prod_2	-0.1287	0.6779	-0.0969	0.7573
d_prod_3	-0.2264	0.4647	-0.2044	0.5139
d_prod_4	-0.3506	0.2793	-0.317	0.3313
d_prod_5	-0.2415	0.4353	-0.217	0.4864
d_prod_6	-0.1712	0.5799	-0.1566	0.6156
d_prod_7	-0.3064	0.3399	-0.265	0.4139
d_dti_1	0.0533	0.6199	0.0437	0.6871
d_dti_2	-0.0252	0.7769	-0.0212	0.8122
d_dti_3	-0.0433	0.5876	-0.0528	0.5151
d_dti_4	0.00446	0.9542	0.015	0.851
d_dti_5	-0.0757	0.3408	-0.0907	0.2581
d_lti_1	0.1338*	0.0923	0.1042	0.1922
d_lti_2	0.1703**	0.0138	0.1518**	0.031
d_lti_3	0.1303**	0.0324	0.1111*	0.0705
d_lti_4	0.1666**	0.0107	0.1666**	0.0126
d_lti_5	0.1102*	0.0844	0.0892	0.1746
d_int_1	-0.1286*	0.061	-0.1387**	0.0444
d_int_2	-0.5179***	<.0001	-0.5148***	<.0001
d_int_3	0.0176	0.8026	-0.00581	0.9346
d_int_4	-0.0936	0.1015	-0.0998*	0.081
d_int_5	-0.4343***	<.0001	-0.4516***	<.0001
d_int_6	-0.2172***	0.0067	-0.2297***	0.0042
d_mst_1	0.00985	0.8623	0.00398	0.9443

Dependent Variable: Recovery Rate				
	OLS O7		OLS O8	
Variables	Estimate	P-Value	Estimate	P-Value
d_mst_2	-5.6E-05	0.9991	-0.00821	0.8734
d_mst_3	0.00247	0.9675	-0.00026	0.9966
d_mst_4	0.0623	0.3922	0.0753	0.303
d_rel_1	0.0747	0.1502	0.0677	0.1946
d_rel_2	0.0865	0.1066	0.0792	0.1417
d_rel_3	0.1676*	0.0837	0.1697*	0.0825
d_rel_4	0.2544**	0.0373	0.2601**	0.0335
d_hst_1	-0.0216	0.652	-0.0267	0.5786
d_hst_2	-0.1004	0.1222	-0.0806	0.2194
d_hst_3	-0.0479	0.3897	-0.0441	0.4268
d_edu_0	0.0232	0.8364	0.0371	0.7462
d_edu_1	-0.0537	0.3591	-0.0303	0.6176
d_edu_2	-0.0363	0.5033	-0.0214	0.6967
d_rea_1	-0.1122	0.3318	-0.1148	0.3222
d_rea_2	-0.1741	0.1452	-0.1774	0.139
d_rea_3	-0.0709	0.5556	-0.0756	0.5306
d_rea_4	-0.0868	0.5877	-0.0836	0.6039
d_rea_5	-0.0514	0.7197	-0.0522	0.7194
d_rea_6	0.1727	0.2127	0.1579	0.2555
SQdist01			0.00744	0.6247
SQdist02			-0.00048	0.9759
SQdist021			0.00324	0.8307
SQdist03			-0.0677***	0.0013
SQdist04			0.0119	0.1356
SQdist05			0.0128	0.4714
SQdist06			-0.0243	0.1784
d_cor7_1			0.0194	0.7315
d_cor7_2			0.079	0.3118
d_cor7_3			-0.00416	0.9363
d_cor7_4			-0.0211	0.6838
d_cor7_5			-0.0384	0.6265
d_12_1	0.0631	0.3275	0.0739	0.2561
d_12_2	0.0866	0.4055	0.1267	0.226
d_12_3	-0.0926	0.4183	-0.082	0.4732
d_12_4	0.0145	0.902	0.0163	0.8902
d_13_1	0.072	0.5094	0.0806	0.4653
d_13_2	0.1579	0.3972	0.1661	0.3764
d_13_3	0.0387	0.8034	0.0498	0.7542
d_14_1	0.0491	0.2896	0.0718	0.1266
d_14_2	0.048	0.5391	0.0585	0.4556
d_14_3	-0.046	0.5486	-0.027	0.7247
d_14_4	-0.00708	0.9082	0.0129	0.8342
d_15_1	0.00549	0.9664	0.0108	0.9339

Dependent Variable: Recovery Rate				
	OLS O7		OLS O8	
Variables	Estimate	P-Value	Estimate	P-Value
d_15_2	-0.1573	0.5155	-0.221	0.3656
d_15_3	0.019	0.9101	0.0237	0.8873
d_16_1	-0.0588	0.1548	-0.0583	0.1672
d_16_2	-0.0197	0.7593	-0.0283	0.6631
d_16_3	-0.028	0.6485	-0.0394	0.5268
d_16_4	0.0428	0.5256	0.0279	0.6807
d_17_1	-0.0345	0.6971	-0.051	0.5708
d_17_2	-0.0564	0.6516	-0.0593	0.6361
d_17_3	-0.0677	0.5164	-0.0852	0.4187
d_18_1	0.1911	0.5686	0.0819	0.8084
d_18_2	0.597	0.1294	0.4955	0.2099
d_18_3	0.3925	0.3156	0.3312	0.4013
d_19_1	-0.0141	0.7741	-0.0203	0.6828
d_19_2	0.1171	0.1596	0.1065	0.2036
d_19_3	0.0139	0.8438	-0.0124	0.8606
d_19_4	-0.0171	0.7789	-0.0347	0.5689
d_20_1	-0.0089	0.8552	-0.00568	0.9072
d_20_2	0.088	0.4969	0.0796	0.5373
d_20_3	0.15*	0.0565	0.1814**	0.0218
d_20_4	-0.1574	0.4708	-0.1367	0.5329
d_21_1	-0.1048	0.5084	-0.0426	0.7924
d_21_2	-0.3063	0.114	-0.254	0.1967
d_21_3	-0.1351	0.4586	-0.0655	0.7251
d_21_4	-0.3716	0.3045	-0.3376	0.3564
***, **, * indicate 1% ,5% and 10% significance levels, respectively.				

Appendix 9. Model Beta O1

Dependent Variable: Recovery Rate						
Beta_O1						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.7976	0.4697	-4.5979	0.9011	-4.5979	0.9011
d_gen	0.1815*	0.0666	-1.4832	0.8392	-1.4832	0.8392
d_age_1	0.3799	0.14	0.5775	0.9743	0.5775	0.9743
d_age_2	0.1672	0.4071	-0.4029	0.9735	-0.4029	0.9735
d_age_3	0.2277	0.2433	-0.467	0.9644	-0.467	0.9644
d_age_4	0.1881	0.3578	0.1013	0.9931	0.1013	0.9931
d_age_5	0.128	0.5297	0.1774	0.9869	0.1774	0.9869
d_sal_1	0.2162	0.4865	0.6493	0.9687	0.6493	0.9687
d_sal_2	0.2419	0.2493	-1.4451	0.9217	-1.4451	0.9217
d_sal_3	0.02534	0.8977	-0.8335	0.9463	-0.8335	0.9463
d_sal_4	0.1554	0.5194	0.4982	0.969	0.4982	0.969
dmp_time_1	-0.3145	0.1528	-0.7958	0.9464	-0.7958	0.9464
dmp_time_2	-0.3324	0.1322	-0.2	0.9868	-0.2	0.9868
dmp_time_3	-0.223	0.3078	-0.3122	0.977	-0.3122	0.977
dmp_time_4	-0.5217**	0.034	0.4512	0.9692	0.4512	0.9692
dmp_time_5	-0.3484	0.231	0.6094	0.961	0.6094	0.961
d_prod_1	-0.1151	0.9004	0.4499	0.9825	0.4499	0.9825
d_prod_2	0.216	0.8126	-0.05171	0.9978	-0.05171	0.9978
d_prod_3	0.09484	0.9168	-0.6371	0.9758	-0.6371	0.9758
d_prod_4	1.2822	0.2202	0.9369	0.971	0.9369	0.971
d_prod_5	0.01314	0.9885	-0.5584	0.9792	-0.5584	0.9792
d_prod_6	-0.02055	0.982	0.6117	0.9725	0.6117	0.9725
d_prod_7	-0.08286	0.9305	0.7868	0.9716	0.7868	0.9716
d_int_1	-0.4069*	0.0884	0.1165	0.993	0.1165	0.993
d_int_2	-1.0025**	0.0122	0.893	0.9572	0.893	0.9572
d_int_3	0.2037	0.4208	0.4974	0.9693	0.4974	0.9693
d_int_4	-0.4102**	0.0404	-1.7993	0.8893	-1.7993	0.8893
d_int_5	-0.6812***	0.004	0.4439	0.9732	0.4439	0.9732
d_int_6	-0.4087	0.1406	0.6344	0.9625	0.6344	0.9625
d_dti_1	0.5586*	0.0994	0.2591	0.9909	0.2591	0.9909
d_dti_2	0.02958	0.9152	-0.6024	0.9746	-0.6024	0.9746
d_dti_3	0.0455	0.8545	-0.5743	0.9705	-0.5743	0.9705
d_dti_4	0.1924	0.4345	0.1532	0.9927	0.1532	0.9927
d_dti_5	0.1188	0.6649	0.6307	0.9687	0.6307	0.9687
d_lti_1	-0.4656*	0.086	0.096	0.9956	0.096	0.9956
d_lti_2	0.09484	0.6754	-0.414	0.9775	-0.414	0.9775
d_lti_3	-0.07286	0.7164	-0.03278	0.9979	-0.03278	0.9979
d_lti_4	-0.2634	0.2086	0.2412	0.983	0.2412	0.983

Dependent Variable: Recovery Rate						
Beta_O1						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_lti_5	-0.1657	0.4161	0.3763	0.9786	0.3763	0.9786
d0	1.6981 (<.0001)					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 10. Model Beta O2

Dependent Variable: Recovery Rate						
Beta O2						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.7454	0.5041	-3.8167	0.9501	-3.8167	0.9501
d_gen	0.162	0.1059	-1.1913	0.9203	-1.1913	0.9203
d_age_1	0.3293	0.217	0.6135	0.9813	0.6135	0.9813
d_age_2	0.1272	0.5646	-0.2476	0.9908	-0.2476	0.9908
d_age_3	0.1937	0.3516	-0.3013	0.9873	-0.3013	0.9873
d_age_4	0.1658	0.4397	0.278	0.9901	0.278	0.9901
d_age_5	0.1193	0.5611	0.364	0.9848	0.364	0.9848
d_sal_1	0.3342	0.2858	0.7451	0.9779	0.7451	0.9779
d_sal_2	0.2814	0.1753	-1.2159	0.9556	-1.2159	0.9556
d_sal_3	0.1339	0.4931	-0.6113	0.9752	-0.6113	0.9752
d_sal_4	0.2499	0.3	0.6422	0.978	0.6422	0.978
d_emp_time_1	-0.2624	0.2233	-0.6477	0.9744	-0.6477	0.9744
d_emp_time_2	-0.2683	0.2116	-0.07185	0.9973	-0.07185	0.9973
d_emp_time_3	-0.115	0.5907	-0.08377	0.9964	-0.08377	0.9964
d_emp_time_4	-0.545**	0.0265	0.525	0.9793	0.525	0.9793
d_emp_time_5	-0.07753	0.7926	0.7601	0.9732	0.7601	0.9732
d_prod_1	-0.1079	0.9075	0.5586	0.9875	0.5586	0.9875
d_prod_2	0.2546	0.7812	0.0938	0.9977	0.0938	0.9977
d_prod_3	0.07308	0.9362	-0.4907	0.9881	-0.4907	0.9881
d_prod_4	0.9647	0.3418	0.948	0.9823	0.948	0.9823
d_prod_5	0.08859	0.9228	-0.4201	0.9907	-0.4201	0.9907
d_prod_6	-0.00369	0.9968	0.7262	0.9816	0.7262	0.9816
d_prod_7	-0.1189	0.9011	0.8553	0.9823	0.8553	0.9823
d_int_1	-0.2505	0.2802	0.33	0.9889	0.33	0.9889
d_int_2	-	0.0472	0.9189	0.9764	0.9189	0.9764
d_int_3	0.29	0.2339	0.602	0.9792	0.602	0.9792
d_int_4	-0.2778	0.1478	-1.5928	0.9393	-1.5928	0.9393
d_int_5	-	0.0257	0.5527	0.9801	0.5527	0.9801
d_int_6	-0.2873	0.2781	0.698	0.9773	0.698	0.9773
d_dti_1	0.4871	0.1563	0.4113	0.9908	0.4113	0.9908
d_dti_2	0.01804	0.9482	-0.3673	0.9896	-0.3673	0.9896
d_dti_3	-0.0241	0.9217	-0.4545	0.9837	-0.4545	0.9837
d_dti_4	0.01338	0.9558	0.238	0.9921	0.238	0.9921
d_dti_5	-0.02062	0.9379	0.7181	0.9767	0.7181	0.9767
d_lti_1	-0.4016	0.1429	0.2564	0.9931	0.2564	0.9931
d_lti_2	0.1636	0.4681	-0.2301	0.992	-0.2301	0.992
d_lti_3	-0.08374	0.6707	0.07105	0.997	0.07105	0.997
d_lti_4	-0.1645	0.4289	0.3451	0.9859	0.3451	0.9859
d_lti_5	-0.0427	0.8341	0.4313	0.9847	0.4313	0.9847
d_edu_0	0.2569	0.4398	0.8525	0.9737	0.8525	0.9737

Dependent Variable: Recovery Rate						
Beta O2						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_edu_1	-0.3322*	0.0705	-0.3599	0.9831	-0.3599	0.9831
d_edu_2	-0.2633	0.1201	-1.6732	0.9122	-1.6732	0.9122
d_hst_1	-0.1791	0.2484	-1.7252	0.9239	-1.7252	0.9239
d_hst_2	-0.2918	0.178	0.6247	0.9765	0.6247	0.9765
d_hst_3	-0.1821	0.3023	0.05495	0.9974	0.05495	0.9974
d_rel_1	0.1713	0.3024	-1.4837	0.9328	-1.4837	0.9328
d_rel_2	0.2489	0.1647	-0.3737	0.9818	-0.3737	0.9818
d_rel_3	0.4817	0.1001	0.859	0.9803	0.859	0.9803
d_rel_4	0.05741	0.8636	0.8859	0.9798	0.8859	0.9798
d_mst_1	0.09061	0.6207	-0.03045	0.9987	-0.03045	0.9987
d_mst_2	0.08191	0.629	-1.0109	0.9583	-1.0109	0.9583
d_mst_3	0.1967	0.3328	0.356	0.9869	0.356	0.9869
d_mst_4	0.0109	0.9607	0.5289	0.9781	0.5289	0.9781
d0	1.7717 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 11. Model Beta O3

Dependent Variable: Recovery Rate						
Beta O3						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.6811	0.5524	-3.7149	0.9695	-3.7149	0.9695
d_gen	0.09754	0.3324	-1.1722	0.9452	-1.1722	0.9452
d_age_1	0.331	0.2235	0.6165	0.9867	0.6165	0.9867
d_age_2	0.1518	0.5063	-0.2423	0.9936	-0.2423	0.9936
d_age_3	0.1567	0.4646	-0.2982	0.9918	-0.2982	0.9918
d_age_4	0.2059	0.3401	0.2823	0.9929	0.2823	0.9929
d_age_5	0.1149	0.5766	0.4014	0.9886	0.4014	0.9886
d_sal_1	0.3927	0.2056	0.7744	0.9841	0.7744	0.9841
d_sal_2	0.299	0.1477	-1.175	0.9704	-1.175	0.9704
d_sal_3	0.04903	0.8013	-0.6011	0.984	-0.6011	0.984
d_sal_4	0.1819	0.4475	0.652	0.9841	0.652	0.9841
d_emp_time_1	-0.1784	0.4136	-0.636	0.9809	-0.636	0.9809
d_emp_time_2	-0.1786	0.4142	-0.06347	0.9982	-0.06347	0.9982
d_emp_time_3	-0.09552	0.6565	-0.06665	0.9978	-0.06665	0.9978
d_emp_time_4	-0.4644*	0.059	0.56	0.9848	0.56	0.9848
d_emp_time_5	-0.02282	0.9369	0.7753	0.9816	0.7753	0.9816
d_prod_1	-0.1099	0.9008	0.5722	0.9909	0.5722	0.9909
d_prod_2	0.2244	0.7964	0.1229	0.9978	0.1229	0.9978
d_prod_3	0.05262	0.9516	-0.4739	0.9919	-0.4739	0.9919
d_prod_4	0.9012	0.3427	0.9488	0.9869	0.9488	0.9869
d_prod_5	0.01296	0.9881	-0.3964	0.9936	-0.3964	0.9936
d_prod_6	-0.07772	0.9286	0.7193	0.9869	0.7193	0.9869
d_prod_7	-0.07131	0.9376	0.8794	0.9873	0.8794	0.9873
d_int_1	-0.1815	0.4266	0.3517	0.992	0.3517	0.992
d_int_2	-0.7075*	0.0679	0.9221	0.9843	0.9221	0.9843
d_int_3	0.3265	0.1769	0.6055	0.986	0.6055	0.986
d_int_4	-0.2346	0.2078	-1.5476	0.9604	-1.5476	0.9604
d_int_5	-	0.0466	0.55	0.9874	0.55	0.9874
d_int_6	-0.2936	0.2603	0.726	0.9844	0.726	0.9844
d_dti_1	0.4907	0.1525	0.4347	0.9927	0.4347	0.9927
d_dti_2	0.04715	0.866	-0.361	0.9926	-0.361	0.9926
d_dti_3	-0.08723	0.7217	-0.4309	0.9881	-0.4309	0.9881
d_dti_4	0.04462	0.8547	0.2654	0.9936	0.2654	0.9936
d_dti_5	-0.04181	0.8754	0.7196	0.9826	0.7196	0.9826
d_lti_1	-0.3257	0.2297	0.2809	0.9944	0.2809	0.9944
d_lti_2	0.1894	0.3939	-0.2059	0.9949	-0.2059	0.9949
d_lti_3	0.00136	0.9944	0.07795	0.9977	0.07795	0.9977
d_lti_4	-0.03812	0.853	0.3593	0.9901	0.3593	0.9901
d_lti_5	0.02187	0.9142	0.436	0.9893	0.436	0.9893
d_edu_0	0.2399	0.4639	0.8591	0.9827	0.8591	0.9827

Dependent Variable: Recovery Rate						
Beta O3						
	Error		Zero		One	
d_edu_1	-0.3224*	0.0732	-0.327	0.989	-0.327	0.989
d_edu_2	-0.2153	0.1944	-1.6529	0.9366	-1.6529	0.9366
d_hst_1	-0.144	0.3447	-1.7064	0.9455	-1.7064	0.9455
d_hst_2	-0.3509	0.1017	0.6326	0.9815	0.6326	0.9815
d_hst_3	-0.1865	0.2813	0.06073	0.9979	0.06073	0.9979
d_rel_1	0.218	0.1918	-1.4325	0.9537	-1.4325	0.9537
d_rel_2	0.2874	0.1082	-0.3717	0.9858	-0.3717	0.9858
d_rel_3	0.5229*	0.0702	0.86	0.9844	0.86	0.9844
d_rel_4	0.13	0.6942	0.8963	0.9868	0.8963	0.9868
d_mst_1	0.1014	0.5777	-0.02165	0.9993	-0.02165	0.9993
d_mst_2	0.08566	0.6118	-0.9936	0.9666	-0.9936	0.9666
d_mst_3	0.167	0.4063	0.3886	0.9899	0.3886	0.9899
d_mst_4	0.0453	0.8366	0.5393	0.9849	0.5393	0.9849
d_rea_1	-0.06139	0.8769	-1.7881	0.9643	-1.7881	0.9643
d_rea_2	-0.2595	0.525	0.0882	0.9982	0.0882	0.9982
d_rea_3	-0.01158	0.9776	0.4807	0.9901	0.4807	0.9901
d_rea_4	0.1774	0.7344	0.9194	0.9862	0.9194	0.9862
d_rea_5	-0.5743	0.2244	0.8591	0.9881	0.8591	0.9881
d_rea_6	0.7689	0.1035	0.8074	0.9855	0.8074	0.9855
d0	1.8092 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 12. Model Beta O4

Dependent Variable: Recovery Rate						
Beta O4						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.7748	0.5008	-3.5713	0.9986	-3.5713	0.9986
d_gen	0.1246	0.2295	-1.1272	0.9966	-1.1272	0.9966
d_age_1	0.3476	0.2188	0.6226	0.9992	0.6226	0.9992
d_age_2	0.1659	0.4909	-0.2081	0.9997	-0.2081	0.9997
d_age_3	0.2161	0.3403	-0.2727	0.9996	-0.2727	0.9996
d_age_4	0.2531	0.2609	0.3069	0.9995	0.3069	0.9995
d_age_5	0.1542	0.4654	0.4337	0.9991	0.4337	0.9991
d_sal_1	0.4353	0.17	0.7965	0.9991	0.7965	0.9991
d_sal_2	0.3152	0.132	-1.1268	0.9983	-1.1268	0.9983
d_sal_3	0.1047	0.5985	-0.5615	0.999	-0.5615	0.999
d_sal_4	0.2193	0.3622	0.6729	0.9991	0.6729	0.9991
d_emp_time_1	-0.1728	0.4376	-0.6107	0.999	-0.6107	0.999
d_emp_time_2	-0.2121	0.3431	-0.04407	0.9999	-0.04407	0.9999
d_emp_time_3	-0.09697	0.6574	-0.00913	1	-0.00913	1
d_emp_time_4	-	0.02	0.5719	0.999	0.5719	0.999
d_emp_time_5	-0.05054	0.8616	0.7975	0.9989	0.7975	0.9989
d_prod_1	-0.1671	0.8464	0.5973	0.9995	0.5973	0.9995
d_prod_2	0.1727	0.8399	0.1446	0.9999	0.1446	0.9999
d_prod_3	0.00945	0.9911	-0.447	0.9996	-0.447	0.9996
d_prod_4	1.1753	0.2169	0.9497	0.9994	0.9497	0.9994
d_prod_5	-0.06044	0.9431	-0.3721	0.9997	-0.3721	0.9997
d_prod_6	-0.04156	0.9611	0.7451	0.9994	0.7451	0.9994
d_prod_7	-0.09737	0.9133	0.8867	0.9993	0.8867	0.9993
d_int_1	-0.3212	0.176	0.3961	0.9995	0.3961	0.9995
d_int_2	-	0.031	0.9244	0.999	0.9244	0.999
d_int_3	0.2626	0.3041	0.6204	0.9992	0.6204	0.9992
d_int_4	-0.3356	0.0891	-1.5159	0.9977	-1.5159	0.9977
d_int_5	-0.6064	0.0081	0.5698	0.9992	0.5698	0.9992
d_int_6	-0.3751	0.1597	0.7484	0.9989	0.7484	0.9989
d_dti_1	0.6114	0.0868	0.4589	0.9996	0.4589	0.9996
d_dti_2	0.06688	0.8145	-0.3214	0.9995	-0.3214	0.9995
d_dti_3	-0.03101	0.902	-0.4091	0.9994	-0.4091	0.9994
d_dti_4	0.07635	0.7603	0.2825	0.9996	0.2825	0.9996
d_dti_5	-0.03917	0.8853	0.7346	0.9989	0.7346	0.9989
d_lti_1	-0.4309	0.1191	0.3079	0.9996	0.3079	0.9996
d_lti_2	0.09341	0.6811	-0.1704	0.9997	-0.1704	0.9997
d_lti_3	-0.06982	0.7241	0.09433	0.9998	0.09433	0.9998
d_lti_4	-0.1198	0.573	0.3708	0.9993	0.3708	0.9993
d_lti_5	-0.09295	0.6504	0.4459	0.9991	0.4459	0.9991
d_edu_0	0.2458	0.4613	0.8612	0.9989	0.8612	0.9989

Dependent Variable: Recovery Rate						
Beta O4						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_edu_1	-0.1352	0.4725	-0.2734	0.9995	-0.2734	0.9995
d_edu_2	-0.105	0.5353	-1.5919	0.9966	-1.5919	0.9966
d_hst_1	-0.1512	0.3328	-1.6683	0.9964	-1.6683	0.9964
d_hst_2	-0.2842	0.2068	0.6728	0.999	0.6728	0.999
d_hst_3	-0.158	0.3688	0.09165	0.9998	0.09165	0.9998
d_rel_1	0.1665	0.3291	-1.3921	0.9975	-1.3921	0.9975
d_rel_2	0.2273	0.2152	-0.3106	0.9995	-0.3106	0.9995
d_rel_3	0.5878**	0.0491	0.8616	0.9995	0.8616	0.9995
d_rel_4	0.06046	0.8576	0.8994	0.999	0.8994	0.999
d_mst_1	0.01887	0.9195	0.0057	1	0.0057	1
d_mst_2	0.01844	0.9164	-0.9497	0.9981	-0.9497	0.9981
d_mst_3	0.1109	0.6007	0.4189	0.9993	0.4189	0.9993
d_mst_4	0.05155	0.8166	0.5581	0.9991	0.5581	0.9991
d_rea_1	0.00186	0.9965	-1.7432	0.9983	-1.7432	0.9983
d_rea_2	-0.1725	0.691	0.1313	0.9999	0.1313	0.9999
d_rea_3	0.03439	0.9371	0.5094	0.9995	0.5094	0.9995
d_rea_4	0.1995	0.7083	0.9245	0.9993	0.9245	0.9993
d_rea_5	-0.3973	0.4292	0.872	0.9993	0.872	0.9993
d_rea_6	0.6969	0.1593	0.8113	0.9994	0.8113	0.9994
corr01	0.02094	0.8708	-2.7634	0.9917	-2.7634	0.9917
corr02	-0.1478	0.3417	-2.9876	0.9918	-2.9876	0.9918
corr021	-0.0161	0.8803	-0.4466	0.999	-0.4466	0.999
corr03	0.1511	0.1603	-1.2389	0.9972	-1.2389	0.9972
corr04	0.05647	0.6224	-0.1583	0.9997	-0.1583	0.9997
corr05	0.4022	0.6468	0.9874	0.9994	0.9874	0.9994
corr06	0.09948	0.3646	-1.4413	0.9967	-1.4413	0.9967
d_cor7_1	-0.2682	0.1617	0.2052	0.9997	0.2052	0.9997
d_cor7_2	-0.0349	0.8894	0.761	0.9989	0.761	0.9989
d_cor7_3	0.0036	0.9837	-0.4486	0.9993	-0.4486	0.9993
d_cor7_4	-0.1096	0.5371	-0.3341	0.9994	-0.3341	0.9994
d_cor7_5	-0.3594	0.1558	0.7358	0.9989	0.7358	0.9989
d0	1.8327 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 13. Model Beta O5

Dependent Variable: Recovery Rate						
Beta O5						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.715	0.5392	-4.0882	0.9893	-4.0882	0.9893
d_gen	0.1037	0.3132	-1.2886	0.9777	-1.2886	0.9777
d_age_1	0.3309	0.2326	0.6012	0.9949	0.6012	0.9949
d_age_2	0.1568	0.5022	-0.3029	0.9967	-0.3029	0.9967
d_age_3	0.1902	0.3869	-0.4446	0.9956	-0.4446	0.9956
d_age_4	0.2739	0.2142	0.2323	0.9975	0.2323	0.9975
d_age_5	0.1648	0.4255	0.36	0.9955	0.36	0.9955
d_sal_1	0.3675	0.2429	0.6868	0.9947	0.6868	0.9947
d_sal_2	0.3381	0.109	-1.3205	0.9871	-1.3205	0.9871
d_sal_3	0.08478	0.6705	-0.6992	0.9925	-0.6992	0.9925
d_sal_4	0.2615	0.2794	0.6441	0.9934	0.6441	0.9934
d_emp_time_1	-0.1782	0.4197	-0.7338	0.9923	-0.7338	0.9923
d_emp_time_2	-0.1836	0.4066	-0.09846	0.999	-0.09846	0.999
d_emp_time_3	-0.08374	0.7009	-0.2055	0.9974	-0.2055	0.9974
d_emp_time_4	-0.485*	0.0547	0.5131	0.9938	0.5131	0.9938
d_emp_time_5	0.09801	0.7396	0.7367	0.9931	0.7367	0.9931
d_prod_1	-0.1092	0.903	0.5415	0.9975	0.5415	0.9975
d_prod_2	0.145	0.8699	0.08561	0.9996	0.08561	0.9996
d_prod_3	0.01053	0.9905	-0.5418	0.9973	-0.5418	0.9973
d_prod_4	1.0255	0.2908	0.9452	0.9961	0.9452	0.9961
d_prod_5	-0.05635	0.9487	-0.5029	0.9974	-0.5029	0.9974
d_prod_6	-0.01408	0.9872	0.6087	0.9969	0.6087	0.9969
d_prod_7	-0.1327	0.886	0.8675	0.9959	0.8675	0.9959
d_int_1	-0.2828	0.218	0.3382	0.9968	0.3382	0.9968
d_int_2	-	0.0449	0.9058	0.993	0.9058	0.993
d_int_3	0.3095	0.2078	0.5317	0.9947	0.5317	0.9947
d_int_4	-0.3022	0.1073	-1.675	0.9817	-1.675	0.9817
d_int_5	-0.555**	0.0124	0.4345	0.9955	0.4345	0.9955
d_int_6	-0.339	0.1941	0.7123	0.9927	0.7123	0.9927
d_dti_1	0.5526	0.1174	0.3917	0.9973	0.3917	0.9973
d_dti_2	-0.00853	0.976	-0.4374	0.9963	-0.4374	0.9963
d_dti_3	-0.09402	0.71	-0.5022	0.9947	-0.5022	0.9947
d_dti_4	-0.0194	0.9385	0.2224	0.998	0.2224	0.998
d_dti_5	-0.1322	0.6282	0.6854	0.9936	0.6854	0.9936
d_lti_1	-0.4047	0.1427	0.2282	0.9982	0.2282	0.9982
d_lti_2	0.1504	0.5102	-0.2655	0.9974	-0.2655	0.9974
d_lti_3	-0.01638	0.9341	0.03734	0.9996	0.03734	0.9996
d_lti_4	-0.09754	0.6468	0.3268	0.9961	0.3268	0.9961
d_lti_5	-0.02923	0.8889	0.3764	0.996	0.3764	0.996
d_edu_0	0.2276	0.495	0.8224	0.9931	0.8224	0.9931

Dependent Variable: Recovery Rate						
Beta O5						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_edu_1	-0.1689	0.3636	-0.5094	0.994	-0.5094	0.994
d_edu_2	-0.1185	0.4822	-1.7883	0.9755	-1.7883	0.9755
d_hst_1	-0.1373	0.3771	-1.8671	0.9756	-1.8671	0.9756
d_hst_2	-0.2426	0.2758	0.6293	0.9939	0.6293	0.9939
d_hst_3	-0.1516	0.3864	-0.09274	0.9987	-0.09274	0.9987
d_rel_1	0.1413	0.407	-1.6089	0.9804	-1.6089	0.9804
d_rel_2	0.2201	0.2329	-0.5164	0.9934	-0.5164	0.9934
d_rel_3	0.5814*	0.0508	0.8556	0.9957	0.8556	0.9957
d_rel_4	0.02149	0.9487	0.8938	0.9936	0.8938	0.9936
d_mst_1	0.01793	0.9233	-0.08865	0.9987	-0.08865	0.9987
d_mst_2	0.03256	0.8519	-1.0666	0.9864	-1.0666	0.9864
d_mst_3	0.1267	0.5438	0.2545	0.9971	0.2545	0.9971
d_mst_4	-0.01845	0.9339	0.477	0.9948	0.477	0.9948
d_rea_1	0.03969	0.9249	-1.9738	0.9868	-1.9738	0.9868
d_rea_2	-0.1616	0.7086	0.013	0.9999	0.013	0.9999
d_rea_3	0.03158	0.942	0.4837	0.9971	0.4837	0.9971
d_rea_4	0.1969	0.7121	0.9147	0.9952	0.9147	0.9952
d_rea_5	-0.3794	0.4493	0.7712	0.9957	0.7712	0.9957
d_rea_6	0.686	0.1615	0.7996	0.9956	0.7996	0.9956
sqdist01	-0.02039	0.6618	0.5203	0.9744	0.5203	0.9744
sqdist02	0.05446	0.2757	0.4236	0.9775	0.4236	0.9775
sqdist021	0.00096	0.9844	0.466	0.9839	0.466	0.9839
sqdist03	-0.3366	0.3426	1.4465	0.9898	1.4465	0.9898
sqdist04	-0.4293	0.2434	1.4666	0.9909	1.4666	0.9909
sqdist05	0.02564	0.6556	0.7338	0.9801	0.7338	0.9801
sqdist06	-	0.0993	0.9573	0.9646	0.9573	0.9646
d_cor7_1	-0.2522	0.1848	0.03003	0.9997	0.03003	0.9997
d_cor7_2	-0.07709	0.7582	0.7465	0.9935	0.7465	0.9935
d_cor7_3	0.02747	0.8746	-0.5951	0.9935	-0.5951	0.9935
d_cor7_4	-0.09093	0.6043	-0.459	0.9952	-0.459	0.9952
d_cor7_5	-0.4139*	0.0952	0.7066	0.9932	0.7066	0.9932
d0	1.8459 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 14. Model Beta O6

Dependent Variable: Recovery Rate						
Beta O6						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.7581	0.5096	-3.7311	0.978	-3.7311	0.978
d_gen	0.1203	0.2434	-1.1695	0.9613	-1.1695	0.9613
d_age_1	0.3684	0.1863	0.6165	0.9903	0.6165	0.9903
d_age_2	0.1627	0.4859	-0.2442	0.9952	-0.2442	0.9952
d_age_3	0.2102	0.3381	-0.3085	0.9936	-0.3085	0.9936
d_age_4	0.2455	0.2634	0.2747	0.9947	0.2747	0.9947
d_age_5	0.1302	0.5274	0.4009	0.991	0.4009	0.991
d_sal_1	0.4326	0.1731	0.7674	0.9888	0.7674	0.9888
d_sal_2	0.358*	0.0879	-1.1792	0.9784	-1.1792	0.9784
d_sal_3	0.1128	0.5709	-0.6043	0.9885	-0.6043	0.9885
d_sal_4	0.2504	0.2999	0.6478	0.9879	0.6478	0.9879
d_emp_time_1	-0.2209	0.3181	-0.6366	0.9855	-0.6366	0.9855
d_emp_time_2	-0.2583	0.2458	-	0.9987	-	0.9987
d_emp_time_3	-0.1413	0.517	-	0.9979	-	0.9979
d_emp_time_4	-0.6088**	0.0143	0.5614	0.988	0.5614	0.988
d_emp_time_5	-0.1153	0.6918	0.7693	0.9859	0.7693	0.9859
d_prod_1	-0.1887	0.8291	0.5694	0.9937	0.5694	0.9937
d_prod_2	0.1593	0.854	0.1172	0.9987	0.1172	0.9987
d_prod_3	-0.01214	0.9887	-0.4719	0.9947	-0.4719	0.9947
d_prod_4	1.2614	0.1927	0.9482	0.9907	0.9482	0.9907
d_prod_5	-0.0526	0.9511	-0.3934	0.9957	-0.3934	0.9957
d_prod_6	-0.04283	0.9604	0.712	0.9918	0.712	0.9918
d_prod_7	-0.08156	0.9281	0.8801	0.9915	0.8801	0.9915
d_int_1	-0.322	0.1677	0.3499	0.9939	0.3499	0.9939
d_int_2	-0.8763**	0.0242	0.9215	0.9888	0.9215	0.9888
d_int_3	0.2888	0.2469	0.6094	0.9895	0.6094	0.9895
d_int_4	-0.3527*	0.0653	-1.5489	0.971	-1.5489	0.971
d_int_5	-	0.0083	0.5425	0.9898	0.5425	0.9898
d_int_6	-0.4077	0.1206	0.7227	0.9872	0.7227	0.9872
d_dti_1	0.5901*	0.0959	0.4317	0.995	0.4317	0.995
d_dti_2	0.03696	0.8969	-0.3651	0.9951	-0.3651	0.9951
d_dti_3	-0.0853	0.7349	-0.4312	0.9924	-0.4312	0.9924
d_dti_4	0.03083	0.9009	0.2668	0.996	0.2668	0.996
d_dti_5	-0.07	0.7957	0.7236	0.9889	0.7236	0.9889
d_lti_1	-0.4516	0.1024	0.2768	0.9961	0.2768	0.9961
d_lti_2	0.09681	0.6712	-0.2077	0.9964	-0.2077	0.9964
d_lti_3	-0.06739	0.7357	0.07724	0.9983	0.07724	0.9983
d_lti_4	-0.1491	0.4849	0.3601	0.9924	0.3601	0.9924
d_lti_5	-0.1106	0.5911	0.4294	0.9922	0.4294	0.9922
d_edu_0	0.3521	0.2885	0.8593	0.9878	0.8593	0.9878

Dependent Variable: Recovery Rate						
Beta O6						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_edu_1	-0.1054	0.5727	-0.3404	0.993	-0.3404	0.993
d_edu_2	-0.05651	0.7382	-1.6629	0.9586	-1.6629	0.9586
d_hst_1	-0.1992	0.2032	-1.714	0.9623	-1.714	0.9623
d_hst_2	-0.3286	0.1426	0.6281	0.9875	0.6281	0.9875
d_hst_3	-0.192	0.2773	0.05063	0.9988	0.05063	0.9988
d_rel_1	0.1551	0.3652	-1.4356	0.9659	-1.4356	0.9659
d_rel_2	0.2441	0.1842	-0.386	0.9893	-0.386	0.9893
d_rel_3	0.6041**	0.0421	0.86	0.9893	0.86	0.9893
d_rel_4	0.1065	0.7529	0.8968	0.9903	0.8968	0.9903
d_mst_1	0.03535	0.8498	-	0.9994	-	0.9994
d_mst_2	0.04789	0.783	-0.9963	0.9785	-0.9963	0.9785
d_mst_3	0.1604	0.4426	0.3787	0.9929	0.3787	0.9929
d_mst_4	0.07829	0.7252	0.5355	0.9895	0.5355	0.9895
d_rea_1	0.00529	0.9901	-1.7876	0.9757	-1.7876	0.9757
d_rea_2	-0.1659	0.704	0.07774	0.999	0.07774	0.999
d_rea_3	0.03589	0.9351	0.4871	0.9935	0.4871	0.9935
d_rea_4	0.1578	0.7697	0.9139	0.9907	0.9139	0.9907
d_rea_5	-0.4457	0.3778	0.8499	0.9918	0.8499	0.9918
d_rea_6	0.7349	0.1402	0.8078	0.9903	0.8078	0.9903
Score	0.3498	0.1157	-0.8758	0.9867	-0.8758	0.9867
d_cor7_1	-0.3024	0.1099	0.1369	0.9973	0.1369	0.9973
d_cor7_2	-0.08461	0.7343	0.7573	0.9881	0.7573	0.9881
d_cor7_3	-0.0108	0.9501	-0.4807	0.9894	-0.4807	0.9894
d_cor7_4	-0.1254	0.473	-0.3626	0.9922	-0.3626	0.9922
d_cor7_5	-0.4482	0.0699	0.718	0.9868	0.718	0.9868
d0	1.828 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 15. Model Beta O6

Dependent Variable: Recovery Rate						
Beta O6						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.7581	0.5096	-3.7311	0.978	-3.7311	0.978
d_gen	0.1203	0.2434	-1.1695	0.9613	-1.1695	0.9613
d_age_1	0.3684	0.1863	0.6165	0.9903	0.6165	0.9903
d_age_2	0.1627	0.4859	-0.2442	0.9952	-0.2442	0.9952
d_age_3	0.2102	0.3381	-0.3085	0.9936	-0.3085	0.9936
d_age_4	0.2455	0.2634	0.2747	0.9947	0.2747	0.9947
d_age_5	0.1302	0.5274	0.4009	0.991	0.4009	0.991
d_sal_1	0.4326	0.1731	0.7674	0.9888	0.7674	0.9888
d_sal_2	0.358*	0.0879	-1.1792	0.9784	-1.1792	0.9784
d_sal_3	0.1128	0.5709	-0.6043	0.9885	-0.6043	0.9885
d_sal_4	0.2504	0.2999	0.6478	0.9879	0.6478	0.9879
d_emp_time_1	-0.2209	0.3181	-0.6366	0.9855	-0.6366	0.9855
d_emp_time_2	-0.2583	0.2458	-	0.9987	-	0.9987
d_emp_time_3	-0.1413	0.517	-	0.9979	-	0.9979
d_emp_time_4	-0.6088**	0.0143	0.5614	0.988	0.5614	0.988
d_emp_time_5	-0.1153	0.6918	0.7693	0.9859	0.7693	0.9859
d_prod_1	-0.1887	0.8291	0.5694	0.9937	0.5694	0.9937
d_prod_2	0.1593	0.854	0.1172	0.9987	0.1172	0.9987
d_prod_3	-0.01214	0.9887	-0.4719	0.9947	-0.4719	0.9947
d_prod_4	1.2614	0.1927	0.9482	0.9907	0.9482	0.9907
d_prod_5	-0.0526	0.9511	-0.3934	0.9957	-0.3934	0.9957
d_prod_6	-0.04283	0.9604	0.712	0.9918	0.712	0.9918
d_prod_7	-0.08156	0.9281	0.8801	0.9915	0.8801	0.9915
d_int_1	-0.322	0.1677	0.3499	0.9939	0.3499	0.9939
d_int_2	-0.8763**	0.0242	0.9215	0.9888	0.9215	0.9888
d_int_3	0.2888	0.2469	0.6094	0.9895	0.6094	0.9895
d_int_4	-0.3527*	0.0653	-1.5489	0.971	-1.5489	0.971
d_int_5	-	0.0083	0.5425	0.9898	0.5425	0.9898
d_int_6	-0.4077	0.1206	0.7227	0.9872	0.7227	0.9872
d_dti_1	0.5901*	0.0959	0.4317	0.995	0.4317	0.995
d_dti_2	0.03696	0.8969	-0.3651	0.9951	-0.3651	0.9951
d_dti_3	-0.0853	0.7349	-0.4312	0.9924	-0.4312	0.9924
d_dti_4	0.03083	0.9009	0.2668	0.996	0.2668	0.996
d_dti_5	-0.07	0.7957	0.7236	0.9889	0.7236	0.9889
d_lti_1	-0.4516	0.1024	0.2768	0.9961	0.2768	0.9961
d_lti_2	0.09681	0.6712	-0.2077	0.9964	-0.2077	0.9964
d_lti_3	-0.06739	0.7357	0.07724	0.9983	0.07724	0.9983
d_lti_4	-0.1491	0.4849	0.3601	0.9924	0.3601	0.9924
d_lti_5	-0.1106	0.5911	0.4294	0.9922	0.4294	0.9922
d_edu_0	0.3521	0.2885	0.8593	0.9878	0.8593	0.9878

Dependent Variable: Recovery Rate						
Beta O6						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_edu_1	-0.1054	0.5727	-0.3404	0.993	-0.3404	0.993
d_edu_2	-0.05651	0.7382	-1.6629	0.9586	-1.6629	0.9586
d_hst_1	-0.1992	0.2032	-1.714	0.9623	-1.714	0.9623
d_hst_2	-0.3286	0.1426	0.6281	0.9875	0.6281	0.9875
d_hst_3	-0.192	0.2773	0.05063	0.9988	0.05063	0.9988
d_rel_1	0.1551	0.3652	-1.4356	0.9659	-1.4356	0.9659
d_rel_2	0.2441	0.1842	-0.386	0.9893	-0.386	0.9893
d_rel_3	0.6041**	0.0421	0.86	0.9893	0.86	0.9893
d_rel_4	0.1065	0.7529	0.8968	0.9903	0.8968	0.9903
d_mst_1	0.03535	0.8498	-	0.9994	-	0.9994
d_mst_2	0.04789	0.783	-0.9963	0.9785	-0.9963	0.9785
d_mst_3	0.1604	0.4426	0.3787	0.9929	0.3787	0.9929
d_mst_4	0.07829	0.7252	0.5355	0.9895	0.5355	0.9895
d_rea_1	0.00529	0.9901	-1.7876	0.9757	-1.7876	0.9757
d_rea_2	-0.1659	0.704	0.07774	0.999	0.07774	0.999
d_rea_3	0.03589	0.9351	0.4871	0.9935	0.4871	0.9935
d_rea_4	0.1578	0.7697	0.9139	0.9907	0.9139	0.9907
d_rea_5	-0.4457	0.3778	0.8499	0.9918	0.8499	0.9918
d_rea_6	0.7349	0.1402	0.8078	0.9903	0.8078	0.9903
Score	0.3498	0.1157	-0.8758	0.9867	-0.8758	0.9867
d_cor7_1	-0.3024	0.1099	0.1369	0.9973	0.1369	0.9973
d_cor7_2	-0.08461	0.7343	0.7573	0.9881	0.7573	0.9881
d_cor7_3	-0.0108	0.9501	-0.4807	0.9894	-0.4807	0.9894
d_cor7_4	-0.1254	0.473	-0.3626	0.9922	-0.3626	0.9922
d_cor7_5	-0.4482	0.0699	0.718	0.9868	0.718	0.9868
d0	1.828 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 16. Model Beta O7

Dependent Variable: Recovery Rate						
Beta O7						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.2838	0.859	-2.4815	0.9991	-2.4815	0.9991
d_gen	0.07173	0.4849	-0.4875	0.9982	-0.4875	0.9982
d_age_1	0.2252	0.4386	0.739	0.9989	0.739	0.9989
d_age_2	0.02686	0.9107	0.1316	0.9998	0.1316	0.9998
d_age_3	0.0298	0.894	0.07963	0.9999	0.07963	0.9999
d_age_4	0.1295	0.5561	0.53	0.9991	0.53	0.9991
d_age_5	-0.04617	0.8303	0.4219	0.9993	0.4219	0.9993
d_sal_1	0.1982	0.5381	0.8606	0.999	0.8606	0.999
d_sal_2	0.375*	0.0743	-0.6547	0.9988	-0.6547	0.9988
d_sal_3	0.07602	0.7004	-0.1311	0.9997	-0.1311	0.9997
d_sal_4	0.1532	0.5307	0.6832	0.9988	0.6832	0.9988
d_emp_time_1	-0.2589	0.2426	-0.124	0.9998	-0.124	0.9998
d_emp_time_2	-0.2574	0.2442	0.2135	0.9996	0.2135	0.9996
d_emp_time_3	-0.1154	0.598	0.1293	0.9998	0.1293	0.9998
d_emp_time_4	-0.5627**	0.0267	0.7161	0.9988	0.7161	0.9988
d_emp_time_5	-0.02353	0.9377	0.8541	0.9987	0.8541	0.9987
d_prod_1	-0.06518	0.9466	0.6919	0.9994	0.6919	0.9994
d_prod_2	0.2546	0.7912	0.3984	0.9996	0.3984	0.9996
d_prod_3	0.04658	0.9612	-0.2307	0.9998	-0.2307	0.9998
d_prod_4	0.7906	0.4456	0.9636	0.9993	0.9636	0.9993
d_prod_5	0.01031	0.9914	-0.0063	1	-0.0063	1
d_prod_6	-0.1682	0.8594	0.8265	0.9992	0.8265	0.9992
d_prod_7	-0.2029	0.8387	0.925	0.9993	0.925	0.9993
d_int_1	-0.1439	0.5413	0.3583	0.9994	0.3583	0.9994
d_int_2	-0.6443	0.1063	0.942	0.9988	0.942	0.9988
d_int_3	0.3843	0.1279	0.7202	0.9987	0.7202	0.9987
d_int_4	-0.1344	0.4906	-0.8237	0.9985	-0.8237	0.9985
d_int_5	-0.4603**	0.0419	0.7153	0.9989	0.7153	0.9989
d_int_6	-0.1895	0.4839	0.8244	0.9989	0.8244	0.9989
d_dti_1	0.5939	0.1041	0.5353	0.9994	0.5353	0.9994
d_dti_2	0.1303	0.6605	0.03402	1	0.03402	1
d_dti_3	-0.0038	0.9885	0.0334	0.9999	0.0334	0.9999
d_dti_4	0.2069	0.4237	0.5006	0.9992	0.5006	0.9992
d_dti_5	0.05134	0.8559	0.6233	0.9989	0.6233	0.9989
d_lti_1	-0.2142	0.4549	0.4233	0.9994	0.4233	0.9994
d_lti_2	0.274	0.242	0.1762	0.9997	0.1762	0.9997
d_lti_3	-0.07864	0.7042	0.3312	0.9993	0.3312	0.9993
d_lti_4	0.07498	0.7237	0.5789	0.9988	0.5789	0.9988
d_lti_5	0.04374	0.8354	0.6136	0.9985	0.6136	0.9985
d_edu_0	0.09222	0.7912	0.908	0.9989	0.908	0.9989

Dependent Variable: Recovery Rate						
Beta O7						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_edu_1	-0.4315**	0.0247	-	0.9999	-	0.9999
d_edu_2	-0.3765**	0.0329	-0.824	0.9975	-0.824	0.9975
d_hst_1	-0.06952	0.6616	-0.8841	0.9976	-0.8841	0.9976
d_hst_2	-0.2291	0.3038	0.4904	0.9988	0.4904	0.9988
d_hst_3	-0.07189	0.6893	0.3878	0.9991	0.3878	0.9991
d_rel_1	0.2664	0.1439	-0.6385	0.9981	-0.6385	0.9981
d_rel_2	0.3947**	0.0388	-0.1279	0.9996	-0.1279	0.9996
d_rel_3	0.5225*	0.0779	0.8782	0.9983	0.8782	0.9983
d_rel_4	0.2493	0.4848	0.8413	0.9988	0.8413	0.9988
d_mst_1	0.02031	0.9153	0.277	0.9995	0.277	0.9995
d_mst_2	0.06294	0.7175	-0.6498	0.9987	-0.6498	0.9987
d_mst_3	0.02426	0.9068	0.5917	0.9988	0.5917	0.9988
d_mst_4	-0.08785	0.7006	0.7062	0.9989	0.7062	0.9989
d_rea_1	-0.1195	0.7735	-1.0248	0.9989	-1.0248	0.9989
d_rea_2	-0.2618	0.5422	0.3796	0.9996	0.3796	0.9996
d_rea_3	-0.1979	0.6443	0.479	0.9995	0.479	0.9995
d_rea_4	0.1178	0.8299	0.9456	0.9991	0.9456	0.9991
d_rea_5	-0.5227	0.295	0.917	0.9993	0.917	0.9993
d_rea_6	0.7069	0.1511	0.8729	0.9992	0.8729	0.9992
q_12_1	0.06129	0.7798	-1.5308	0.9973	-1.5308	0.9973
q_12_2	0.2388	0.4928	0.8595	0.9988	0.8595	0.9988
q_12_3	-0.3947	0.3144	0.4915	0.9994	0.4915	0.9994
q_12_4	-0.1464	0.7083	0.9272	0.9988	0.9272	0.9988
q_13_1	0.2879	0.402	-1.801	0.9972	-1.801	0.9972
q_13_2	0.02505	0.9659	0.9185	0.9991	0.9185	0.9991
q_13_3	-0.08593	0.8574	0.4972	0.9994	0.4972	0.9994
q_13_4	0	.	1	<.0001	1	<.0001
q_14_1	0.1557	0.3073	0.3509	0.9992	0.3509	0.9992
q_14_2	-0.00794	0.9748	0.7975	0.9984	0.7975	0.9984
q_14_3	-0.1142	0.6603	0.6407	0.9987	0.6407	0.9987
q_14_4	0.06643	0.7182	0.5463	0.9988	0.5463	0.9988
q_15_1	0.1523	0.6915	-1.923	0.9973	-1.923	0.9973
q_15_2	0.1466	0.8624	0.9587	0.9992	0.9587	0.9992
q_15_3	0.418	0.4248	0.5385	0.9994	0.5385	0.9994
q_15_4	0	.	1	<.0001	1	<.0001
q_16_1	-0.1655	0.2207	-0.517	0.9987	-0.517	0.9987
q_16_2	-0.08326	0.7183	0.6824	0.9987	0.6824	0.9987
q_16_3	-0.3648*	0.0635	0.2665	0.9994	0.2665	0.9994
q_16_4	0.06117	0.7856	0.7811	0.9984	0.7811	0.9984
q_17_1	-0.1095	0.714	-1.5888	0.9976	-1.5888	0.9976
q_17_2	0.2215	0.6248	0.8803	0.9989	0.8803	0.9989
q_17_3	-0.01236	0.9728	0.3292	0.9995	0.3292	0.9995
q_17_4	0	.	1	<.0001	1	<.0001

Dependent Variable: Recovery Rate						
Beta O7						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
q_18_1	-0.142	0.8974	-2.4233	0.9984	-2.4233	0.9984
q_18_2	0.1815	0.8886	0.9682	0.9995	0.9682	0.9995
q_18_3	0.3123	0.806	0.982	0.9993	0.982	0.9993
q_18_4	0	.	1	<.0001	1	<.0001
q_19_1	-0.0699	0.6589	0.3912	0.9991	0.3912	0.9991
q_19_2	1.1614***	0.0009	0.8791	0.9986	0.8791	0.9986
q_19_3	-0.1206	0.6069	0.6083	0.9989	0.6083	0.9989
q_19_4	-0.03958	0.8325	0.5049	0.999	0.5049	0.999
q_20_1	0.2413	0.1191	-1.332	0.997	-1.332	0.997
q_20_2	-0.05987	0.8873	0.9285	0.9986	0.9285	0.9986
q_20_3	0.7571***	0.0022	0.3867	0.9992	0.3867	0.9992
q_20_4	-0.5769	0.3599	0.9514	0.9992	0.9514	0.9992
q_21_1	0.1321	0.8483	-2.2368	0.9978	-2.2368	0.9978
q_21_2	0.2947	0.7086	0.9443	0.9992	0.9443	0.9992
q_21_3	-0.2967	0.6967	0.8447	0.9993	0.8447	0.9993
q_21_4	-0.3439	0.761	0.9917	0.9995	0.9917	0.9995
d0	1.9227 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 17. Model Beta O8

Dependent Variable: Recovery Rate						
Beta O8						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.2626	0.8658	-2.5132	0.9996	-2.5132	0.9996
d_gen	0.08631	0.4062	-0.5238	0.999	-0.5238	0.999
d_age_1	0.1569	0.5949	0.7269	0.9995	0.7269	0.9995
d_age_2	0.01747	0.943	0.09769	0.9999	0.09769	0.9999
d_age_3	0.06653	0.7727	0.0236	1	0.0236	1
d_age_4	0.2111	0.3443	0.504	0.9995	0.504	0.9995
d_age_5	0.01774	0.9344	0.5132	0.9995	0.5132	0.9995
d_sal_1	0.1497	0.645	0.8504	0.9994	0.8504	0.9994
d_sal_2	0.4061*	0.0569	-0.6788	0.9994	-0.6788	0.9994
d_sal_3	0.09988	0.6189	-0.152	0.9999	-0.152	0.9999
d_sal_4	0.1776	0.4648	0.7314	0.9995	0.7314	0.9995
d_emp_time_1	-0.217	0.3259	-0.1808	0.9998	-0.1808	0.9998
d_emp_time_2	-0.224	0.3111	0.1714	0.9998	0.1714	0.9998
d_emp_time_3	-0.03347	0.8794	0.2043	0.9998	0.2043	0.9998
d_emp_time_4	-0.5255**	0.0412	0.6898	0.9993	0.6898	0.9993
d_emp_time_5	0.2214	0.4803	0.8371	0.9993	0.8371	0.9993
d_prod_1	-0.05847	0.9507	0.7053	0.9997	0.7053	0.9997
d_prod_2	0.2166	0.8159	0.3983	0.9998	0.3983	0.9998
d_prod_3	0.0181	0.9845	-0.2202	0.9999	-0.2202	0.9999
d_prod_4	0.7163	0.477	0.9654	0.9996	0.9654	0.9996
d_prod_5	-0.03094	0.9732	-	1	-	1
d_prod_6	-0.1362	0.8814	0.8002	0.9996	0.8002	0.9996
d_prod_7	-0.2106	0.8276	0.9113	0.9996	0.9113	0.9996
d_int_1	-0.1559	0.5043	0.4788	0.9996	0.4788	0.9996
d_int_2	-0.5529	0.1712	0.9357	0.9993	0.9357	0.9993
d_int_3	0.3895	0.1196	0.6796	0.9992	0.6796	0.9992
d_int_4	-0.1066	0.5875	-0.8951	0.9989	-0.8951	0.9989
d_int_5	-0.494**	0.0284	0.6906	0.9993	0.6906	0.9993
d_int_6	-0.1649	0.5407	0.8202	0.9992	0.8202	0.9992
d_dti_1	0.5286	0.1536	0.5728	0.9996	0.5728	0.9996
d_dti_2	0.09748	0.7415	0.00924	1	0.00924	1
d_dti_3	-0.07852	0.7689	-	1	-	1
d_dti_4	0.05742	0.8282	0.4254	0.9996	0.4254	0.9996
d_dti_5	-0.1158	0.6858	0.7137	0.9992	0.7137	0.9992
d_lti_1	-0.2962	0.3059	0.4561	0.9996	0.4561	0.9996
d_lti_2	0.2057	0.3882	0.1354	0.9999	0.1354	0.9999
d_lti_3	-0.09787	0.6392	0.335	0.9996	0.335	0.9996
d_lti_4	0.03666	0.8674	0.5616	0.9994	0.5616	0.9994
d_lti_5	0.03757	0.8616	0.5347	0.9993	0.5347	0.9993
d_edu_0	0.1383	0.6989	0.8899	0.9994	0.8899	0.9994

Dependent Variable: Recovery Rate						
Beta O8						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
d_edu_1	-0.2955	0.1353	-	1	-	1
d_edu_2	-0.2914	0.1023	-0.9501	0.9986	-0.9501	0.9986
d_hst_1	-0.01213	0.939	-0.9721	0.9986	-0.9721	0.9986
d_hst_2	-0.1616	0.4774	0.6446	0.9993	0.6446	0.9993
d_hst_3	-0.01436	0.9359	0.3346	0.9995	0.3346	0.9995
d_rel_1	0.1329	0.4727	-0.7022	0.999	-0.7022	0.999
d_rel_2	0.2495	0.1956	-	0.9999	-	0.9999
d_rel_3	0.5316*	0.078	0.8842	0.9993	0.8842	0.9993
d_rel_4	0.1113	0.7577	0.8893	0.9993	0.8893	0.9993
d_mst_1	-0.0218	0.9106	0.2335	0.9998	0.2335	0.9998
d_mst_2	-0.03196	0.8569	-0.5621	0.9994	-0.5621	0.9994
d_mst_3	-0.0296	0.8901	0.5696	0.9994	0.5696	0.9994
d_mst_4	-0.1338	0.5651	0.6745	0.9994	0.6745	0.9994
d_rea_1	-0.06516	0.8811	-1.0843	0.9995	-1.0843	0.9995
d_rea_2	-0.1901	0.6722	0.315	0.9999	0.315	0.9999
d_rea_3	-0.2227	0.6168	0.5883	0.9997	0.5883	0.9997
d_rea_4	0.128	0.8205	0.948	0.9996	0.948	0.9996
d_rea_5	-0.3273	0.5302	0.9076	0.9996	0.9076	0.9996
d_rea_6	0.6488	0.2004	0.8692	0.9997	0.8692	0.9997
sqdist01	-0.02933	0.5457	0.6716	0.9964	0.6716	0.9964
sqdist02	0.04559	0.383	0.5795	0.997	0.5795	0.997
sqdist021	0.0134	0.7908	0.5016	0.998	0.5016	0.998
sqdist03	-0.5237	0.1934	1.3127	0.9991	1.3127	0.9991
sqdist04	-0.2641	0.5121	1.3405	0.9993	1.3405	0.9993
sqdist05	0.08131	0.1716	0.8006	0.9962	0.8006	0.9962
sqdist06	-	0.0097	1.0059	0.9961	1.0059	0.9961
d_cor7_1	-0.1868	0.3485	0.427	0.9996	0.427	0.9996
d_cor7_2	0.00626	0.9804	0.8359	0.9994	0.8359	0.9994
d_cor7_3	0.05016	0.7831	-0.1223	0.9999	-0.1223	0.9999
d_cor7_4	-0.09248	0.6105	-0.1199	0.9999	-0.1199	0.9999
d_cor7_5	-0.511**	0.0485	0.8	0.9993	0.8	0.9993
q_12_1	0.07312	0.7476	-1.7024	0.9986	-1.7024	0.9986
q_12_2	0.2269	0.516	0.8653	0.9994	0.8653	0.9994
q_12_3	-0.3718	0.3389	0.647	0.9996	0.647	0.9996
q_12_4	-0.07156	0.8598	0.9268	0.9995	0.9268	0.9995
q_13_1	0.3455	0.3216	-1.9945	0.9987	-1.9945	0.9987
q_13_2	-0.04014	0.9462	0.9339	0.9997	0.9339	0.9997
q_13_3	-0.08375	0.8634	0.6497	0.9996	0.6497	0.9996
q_13_4	0	.	1	<.0001	1	<.0001
q_14_1	0.1546	0.3173	0.2883	0.9996	0.2883	0.9996
q_14_2	0.06548	0.796	0.7865	0.9993	0.7865	0.9993
q_14_3	-0.09123	0.7244	0.7382	0.9993	0.7382	0.9993
q_14_4	0.05301	0.7726	0.5301	0.9994	0.5301	0.9994

Dependent Variable: Recovery Rate						
Beta O8						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
q_15_1	0.1821	0.6378	-2.0995	0.9985	-2.0995	0.9985
q_15_2	0.1008	0.9047	0.9774	0.9996	0.9774	0.9996
q_15_3	0.2858	0.5843	0.6662	0.9996	0.6662	0.9996
q_15_4	0	.	1	<.0001	1	<.0001
q_16_1	-0.1651	0.2324	-0.6149	0.9989	-0.6149	0.9989
q_16_2	-0.1697	0.4677	0.7001	0.9994	0.7001	0.9994
q_16_3	-0.3642*	0.0676	0.3478	0.9996	0.3478	0.9996
q_16_4	0.01603	0.9444	0.7703	0.9992	0.7703	0.9992
q_17_1	-0.1245	0.6831	-1.7499	0.9983	-1.7499	0.9983
q_17_2	0.2406	0.5992	0.8875	0.9994	0.8875	0.9994
q_17_3	-0.01874	0.9591	0.4683	0.9996	0.4683	0.9996
q_17_4	0	.	1	<.0001	1	<.0001
q_18_1	-0.1462	0.8985	-2.4548	0.9992	-2.4548	0.9992
q_18_2	0.1136	0.9319	0.9735	0.9997	0.9735	0.9997
q_18_3	0.354	0.787	0.9767	0.9997	0.9767	0.9997
q_18_4	0	.	1	<.0001	1	<.0001
q_19_1	-0.07445	0.6431	0.3517	0.9995	0.3517	0.9995
q_19_2	1.0571***	0.0026	0.8765	0.9995	0.8765	0.9995
q_19_3	-0.2235	0.341	0.7042	0.9993	0.7042	0.9993
q_19_4	-0.07298	0.6972	0.4777	0.9994	0.4777	0.9994
q_20_1	0.2212	0.1558	-1.448	0.998	-1.448	0.998
q_20_2	-0.08858	0.8324	0.9322	0.9994	0.9322	0.9994
q_20_3	0.8611***	0.0006	0.4587	0.9995	0.4587	0.9995
q_20_4	-0.4765	0.4767	0.9669	0.9996	0.9669	0.9996
q_21_1	0.123	0.8665	-2.2568	0.999	-2.2568	0.999
q_21_2	0.2539	0.7583	0.9468	0.9997	0.9468	0.9997
q_21_3	-0.1914	0.8094	0.8319	0.9997	0.8319	0.9997
q_21_4	-0.2423	0.8356	0.9914	0.9997	0.9914	0.9997
d0	1.9703 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 18. Model OLS T1 and OLS T2

Dependent Variable: Recovery Rate				
	OLS T1		OLS T2	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	0.3743***	<.0001	0.3391***	0.0002
d_Gen_1	-0.0168	0.5931	-0.0211	0.5032
d_Age_1	-0.0443	0.2857	-0.0203	0.6456
d_Age_2	-0.0363	0.34	-0.026	0.5006
d_emp_1	-0.0139	0.7711	-0.0182	0.7064
d_emp_2	-0.0171	0.7259	-0.0227	0.6447
d_emp_3	-0.00048	0.9918	0.00059	0.9899
d_sal_1	0.0612	0.1167	0.0546	0.1635
d_sal_2	0.0169	0.7395	0.0195	0.7027
d_sal_3	0.0803	0.1718	0.0801	0.1748
d_prod_1	-0.1078**	0.0287	-0.1099**	0.0263
d_prod_2	-0.1061***	0.0065	-0.1093***	0.0053
d_prod_3	-0.00128	0.9758	-0.00414	0.9222
d_dti_1	0.101	0.2877	0.0907	0.3425
d_dti_2	0.00415	0.9551	-0.00128	0.9863
d_dti_3	0.00528	0.9325	0.00416	0.947
d_dti_4	0.0567	0.3594	0.0503	0.4191
d_lti_1	0.0448	0.5771	0.0549	0.5002
d_lti_2	0.135*	0.0531	0.1416**	0.0452
d_lti_3	0.1195*	0.0533	0.1176*	0.0598
d_lti_4	0.0756	0.1876	0.0843	0.1496
d_inter_1	0.1401***	0.0011	0.1479***	0.0006
d_inter_2	0.1287***	0.0003	0.1325***	0.0002
d_mst_1			-0.015	0.6664
d_rel_1			0.0526	0.2197
d_rel_2			0.0748	0.1116
d_hst_1			0.0356	0.2737
d_edu_1			-0.0462	0.3268
***, **, * indicate 1% ,5% and 10% significance levels, respectively.				

Appendix 19. Model OLS T3 and OLS T4

Dependent Variable: Recovery Rate				
	OLS T3		OLS T4	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	0.3795***	<.0001	0.3809	0.133
d_Gen_1	-0.017	0.5927	-0.0195	0.5514
d_Age_1	-0.0179	0.6891	-0.0215	0.6414
d_Age_2	-0.0315	0.4199	-0.0273	0.4952
d_emp_1	-0.0117	0.8075	-0.00847	0.8626
d_emp_2	-0.0159	0.7451	-0.00804	0.8725
d_emp_3	2.87E-06	1	0.0047	0.9213
d_sal_1	0.0568	0.1452	0.0627	0.1154
d_sal_2	0.0155	0.761	0.0156	0.7633
d_sal_3	0.0963	0.1031	0.0979	0.1068
d_prod_1	-0.1087**	0.0273	-0.1056**	0.036
d_prod_2	-0.1111***	0.0044	-0.112***	0.0056
d_prod_3	0.00119	0.9776	0.00255	0.9538
d_dti_1	0.1017	0.2879	0.1017	0.3005
d_dti_2	-0.00191	0.9795	-0.00141	0.9852
d_dti_3	0.00434	0.9448	0.00023	0.9972
d_dti_4	0.0436	0.4841	0.0424	0.5074
d_lti_1	0.0573	0.4801	0.0579	0.4848
d_lti_2	0.1447**	0.0406	0.1491**	0.0392
d_lti_3	0.1214*	0.0513	0.1232*	0.0535
d_lti_4	0.0839	0.1492	0.0871	0.1446
d_inter_1	0.144***	0.0009	0.1475***	0.0008
d_inter_2	0.1291***	0.0003	0.1335***	0.0003
d_mst_1	-0.00394	0.9099	-0.00378	0.9149
d_rel_1	0.056	0.1901	0.0484	0.2698
d_rel_2	0.0767	0.1028	0.0665	0.1667
d_hst_1	0.028	0.3904	0.0219	0.5168
d_edu_1	-0.0453	0.3364	-0.0477	0.3272
d_rea_1	-0.0576	0.1463	-0.0589	0.1502
d_rea_2	-0.1185**	0.0124	-0.1226**	0.0115
d_cor1			0.0342	0.4359
d_cor2			-0.0364	0.4802
d_cor21			-0.0117	0.7319
d_cor3			-0.0221	0.5249
d_cor4			0.0125	0.7317
d_cor5			0.0128	0.9566
d_cor6			-0.00315	0.9315
d_cor7_1			0.0142	0.8133
d_cor7_2			0.0482	0.5665
d_cor7_3			0.0101	0.8536
d_cor7_4			-0.0384	0.4849

Dependent Variable: Recovery Rate				
	OLS T3		OLS T4	
Variables	Estimate	P-Value	Estimate	P-Value
d_cor7_5			0.0425	0.6043
***, **, * indicate 1% ,5% and 10% significance levels, respectively.				

Appendix 20. Model OLS T5 and OLS T6

Dependent Variable: Recovery Rate				
	OLS T5		OLS T6	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	0.3125**	0.0386	0.3557***	0.001
d_Gen_1	-0.0499	0.2127	-0.0161	0.617
d_Age_1	-0.1026*	0.0699	-0.0211	0.6441
d_Age_2	-0.1242**	0.0106	-0.0276	0.4865
d_emp_1	0.0383	0.5183	-0.00585	0.9041
d_emp_2	-0.00894	0.8869	-0.00848	0.8641
d_emp_3	-0.016	0.7819	0.00583	0.9018
d_sal_1	0.0667	0.1486	0.0615	0.1185
d_sal_2	0.00939	0.8836	0.0152	0.7671
d_sal_3	0.2464***	0.0026	0.0984	0.1014
d_prod_1	-0.1054	0.1044	-0.1036**	0.0371
d_prod_2	-0.1079**	0.0305	-0.1066***	0.0074
d_prod_3	-0.0412	0.4361	0.0051	0.9068
d_dti_1	0.3213***	0.0099	0.1057	0.2765
d_dti_2	0.0353	0.6742	0.00297	0.9683
d_dti_3	0.0663	0.3499	0.00652	0.9183
d_dti_4	0.1372*	0.0673	0.0443	0.4839
d_lti_1	-0.00783	0.9417	0.0582	0.4771
d_lti_2	0.1143	0.1851	0.1475**	0.0389
d_lti_3	0.0586	0.4407	0.1244**	0.0487
d_lti_4	0.00094	0.9892	0.0907	0.1244
d_inter_1	0.143**	0.0109	0.1452***	0.0009
d_inter_2	0.1005**	0.024	0.1345***	0.0002
d_mst_1	-0.0295	0.4963	-0.00476	0.8918
d_rel_1	0.0529	0.3282	0.0495	0.2554
d_rel_2	0.0674	0.2603	0.0701	0.1416
d_hst_1	0.00899	0.8231	0.0245	0.4586
d_edu_1	0.0207	0.7417	-0.0473	0.3278
d_rea_1	-0.0417	0.3941	-0.0561	0.1626
d_rea_2	-0.0905	0.1285	-0.1216**	0.0114
cdist01	-0.017	0.7467		

Dependent Variable: Recovery Rate				
	OLS T5		OLS T6	
Variables	Estimate	P-Value	Estimate	P-Value
cdist02	0.0107	0.8518		
cdist021	-0.0209	0.1303		
cdist03	1.1E-05	0.9982		
cdist04	0.0003*	0.0678		
cdist05	-0.00038	0.3566		
cdist06	-4.5E-05	0.7138		
score			0.0405	0.5682
d_cor7_1	-0.0199	0.7763	0.00848	0.8864
d_cor7_2	0.0341	0.7465	0.0451	0.5853
d_cor7_3	0.0136	0.8302	0.00743	0.89
d_cor7_4	-0.0191	0.7646	-0.0418	0.4374
d_cor7_5	0.0748	0.4525	0.0347	0.6673
***, **, * indicate 1% ,5% and 10% significance levels, respectively.				

Appendix 21. Model OLS T7 and OLS T8

Dependent Variable: Recovery Rate				
	OLS T7		OLS T8	
Variables	Estimate	P-Value	Estimate	P-Value
Intercept	0.6599**	0.0186	0.6542**	0.0275
d_Gen_1	-0.0122	0.7059	-0.0197	0.5568
d_Age_1	-0.00655	0.8867	-0.0101	0.8297
d_Age_2	-0.0279	0.4825	-0.0256	0.5263
d_emp_1	-0.0258	0.5985	-0.0217	0.6615
d_emp_2	-0.0266	0.5926	-0.026	0.6062
d_emp_3	-0.0106	0.8239	-0.00628	0.8968
d_sal_1	0.0466	0.2447	0.0555	0.1734
d_sal_2	0.0205	0.6968	0.0146	0.784
d_sal_3	0.1011*	0.0933	0.1075*	0.0805
d_prod_1	-0.1126**	0.0262	-0.1136**	0.0267
d_prod_2	-0.1104***	0.0059	-0.1184***	0.0043
d_prod_3	-0.0101	0.8153	-0.0135	0.7606
d_dti_1	0.1217	0.219	0.1054	0.296
d_dti_2	0.00752	0.9212	0.00127	0.9868
d_dti_3	0.00898	0.8889	-0.00658	0.9202
d_dti_4	0.0673	0.2935	0.0619	0.3484
d_lti_1	0.0514	0.5374	0.0544	0.5227
d_lti_2	0.1377*	0.0571	0.1513**	0.0408
d_lti_3	0.1108*	0.0818	0.1139*	0.0794
d_lti_4	0.0865	0.1464	0.0948	0.1181
d_inter_1	0.1485***	0.0007	0.1584***	0.0004
d_inter_2	0.141***	0.0001	0.1515***	<.0001
d_mst_1	-0.00071	0.9841	-0.00071	0.9843
d_rel_1	0.0617	0.1716	0.0553	0.2267
d_rel_2	0.0791	0.1077	0.0665	0.1826
d_hst_1	0.0285	0.3918	0.0135	0.6947
d_edu_1	-0.0371	0.4483	-0.0335	0.504
d_rea_1	-0.0683*	0.092	-0.0617	0.1374
d_rea_2	-0.1198**	0.0144	-0.1125**	0.0248
cdist01			0.0439	0.3279
cdist02			-0.0648	0.2145
cdist021			-0.00056	0.935
cdist03			-4.4E-05	0.9726
cdist04			0.00014	0.1472
cdist05			0.00034	0.276
cdist06			-0.00021*	0.0516
d_cor7_1			-0.00448	0.9425
d_cor7_2			0.0284	0.7405
d_cor7_3			-0.0214	0.7068
d_cor7_4			-0.062	0.27

Dependent Variable: Recovery Rate				
	OLS T7		OLS T8	
Variables	Estimate	P-Value	Estimate	P-Value
d_cor7_5			-0.0294	0.7365
dq_12_1	0.15	0.1394	0.1686	0.1012
dq_12_2	0.0848	0.4644	0.0978	0.404
dq_13_1	-0.0612	0.613	-0.0842	0.4984
dq_13_2	0.1535	0.6881	0.115	0.766
dq_14_1	0.0615	0.1422	0.056	0.1884
dq_14_2	-0.037	0.4001	-0.0388	0.3889
dq_15_1	-0.00156	0.9895	-0.00603	0.9599
dq_15_2	-0.11	0.5402	-0.1294	0.4755
dq_16_1	-0.0112	0.8478	-0.0174	0.7663
dq_16_2	0.0584	0.346	0.055	0.3812
dq_17_1	0.0142	0.8315	0.0138	0.8394
dq_17_2	0.0871	0.439	0.1053	0.3598
dq_18_1	-0.2961	0.2072	-0.322	0.1784
dq_18_2	-0.2629	0.52	-0.3114	0.4586
dq_19_1	-0.0409	0.5512	-0.021	0.7656
dq_19_2	-0.028	0.6484	-0.0136	0.8271
dq_20_1	-0.1323	0.0708	-0.1633	0.0287
dq_20_2	-0.1836	0.0273	-0.221	0.0092
dq_21_1	0.0498	0.4625	0.0729	0.2896
dq_21_2	-0.2135	0.5314	-0.1986	0.566
***, **, * indicate 1% ,5% and 10% significance levels, respectively.				

Appendix 22. Model Beta T1

Dependent Variable: Recovery Rate						
Beta T1						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-Value	Estimate	P-Value
Intercept	0.2444	0.3483	-4.5293	0.9785	-4.5293	0.9785
d_gen_1	0.1727*	0.0777	-1.5569	0.9266	-1.5569	0.9266
d_age_1	0.02806	0.8477	-0.8187	0.9722	-0.8187	0.9722
d_age_2	0.04178	0.7627	-0.581	0.9725	-0.581	0.9725
d_age_3	0.1061	0.4878	0.1241	0.9948	0.1241	0.9948
d_emp_1	0.00288	0.9845	-1.0563	0.9948	-1.0563	0.9948
d_emp_2	-0.0175	0.9087	-2.862	0.9797	-2.862	0.9797
d_emp_3	0.1107	0.4581	-3.605	0.9124	-3.605	0.9124
d_sal_1	0.2488	0.0437	-0.9525	0.9517	-0.9525	0.9517
d_sal_2	0.06717	0.6716	-0.286	0.9889	-0.286	0.9889
d_sal_3	0.08795	0.6261	-0.2446	0.9879	-0.2446	0.9879
d_prod_1	0.02062	0.8932	0.2627	0.9844	0.2627	0.9844
d_prod_2	0.04283	0.7339	-0.7704	0.9662	-0.7704	0.9662
d_prod_3	0.1479	0.2602	-0.3625	0.9836	-0.3625	0.9836
d_dti_1	0.1918	0.5312	0.2715	0.9988	0.2715	0.9988
d_dti_2	-0.06556	0.7731	-0.6298	0.9971	-0.6298	0.9971
d_dti_3	-0.03591	0.8493	-0.6685	0.9968	-0.6685	0.9968
d_dti_4	0.09826	0.6186	0.1191	0.9994	0.1191	0.9994
d_lti_1	-0.1226	0.6506	0.09011	0.9984	0.09011	0.9984
d_lti_2	0.2847	0.2039	-0.4091	0.993	-0.4091	0.993
d_lti_3	0.03006	0.8795	-0.1075	0.9983	-0.1075	0.9983
d_lti_4	-0.07842	0.6673	-0.4243	0.9913	-0.4243	0.9913
d_inter_1	0.2297	0.114	-0.3726	0.9842	-0.3726	0.9842
d_inter_2	0.04198	0.7217	-1.9696	0.924	-1.9696	0.924
d0	1.566 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 23. Model Beta T2

Dependent Variable: Recovery Rate						
Beta T2						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.5242*	0.067	-3.2051	0.9665	-3.2051	0.9665
d_gen_1	0.1321	0.1774	-0.9363	0.9545	-0.9363	0.9545
d_age_1	0.05302	0.7255	-0.3896	0.9876	-0.3896	0.9876
d_age_2	0.0891	0.521	-0.2012	0.9897	-0.2012	0.9897
d_age_3	0.2102	0.1692	0.3441	0.9877	0.3441	0.9877
d_emp_1	0.2897**	0.0177	-0.5673	0.9905	-0.5673	0.9905
d_emp_2	-0.01702	0.9158	-1.9311	0.9593	-1.9311	0.9593
d_emp_3	0.06473	0.7254	-2.4963	0.9258	-2.4963	0.9258
d_sal_1	-0.03396	0.8192	-0.4917	0.9807	-0.4917	0.9807
d_sal_2	-0.01629	0.9151	0.02469	0.9989	0.02469	0.9989
d_sal_3	0.1338	0.3721	0.06369	0.9978	0.06369	0.9978
d_prod_1	0.02165	0.8869	0.4397	0.9802	0.4397	0.9802
d_prod_2	0.02969	0.8137	-0.3492	0.9848	-0.3492	0.9848
d_prod_3	0.1249	0.3361	-0.0333	0.9986	-0.0333	0.9986
d_dti_1	0.4343	0.1577	0.4444	0.9959	0.4444	0.9959
d_dti_2	-0.1078	0.6299	-0.2315	0.9977	-0.2315	0.9977
d_dti_3	-0.04248	0.8187	-0.2661	0.9971	-0.2661	0.9971
d_dti_4	0.08695	0.653	0.3253	0.9966	0.3253	0.9966
d_lti_1	-0.2603	0.3342	0.3077	0.9946	0.3077	0.9946
d_lti_2	0.454**	0.043	-0.06655	0.9986	-0.06655	0.9986
d_lti_3	0.0942	0.6311	0.1546	0.9971	0.1546	0.9971
d_lti_4	-0.00179	0.9922	-0.07333	0.9984	-0.07333	0.9984
d_inter_1	0.3001	0.0394	-0.04282	0.9986	-0.04282	0.9986
d_inter_2	0.06912	0.5535	-1.2585	0.9599	-1.2585	0.9599
d_mst_1	0.06272	0.5604	-0.4578	0.9819	-0.4578	0.9819
d_rel_1	0.1049	0.4269	-1.1276	0.9464	-1.1276	0.9464
d_rel_2	0.1631	0.2713	-0.2566	0.9897	-0.2566	0.9897
d_hst_1	-0.06505	0.5168	-1.7062	0.9093	-1.7062	0.9093
d_edu_1	-	0.001	-2.5195	0.8873	-2.5195	0.8873
d0	1.5899 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 24. Model Beta T3

Dependent Variable: Recovery Rate						
Beta T3						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.5389*	0.0667	-3.2228	0.9801	-3.2228	0.9801
d_gen_1	0.1615*	0.0927	-0.9478	0.9734	-0.9478	0.9734
d_age_1	0.08956	0.5504	-0.4125	0.9918	-0.4125	0.9918
d_age_2	0.08389	0.5417	-0.2066	0.9943	-0.2066	0.9943
d_age_3	0.1849	0.2141	0.3465	0.9926	0.3465	0.9926
d_emp_1	0.2443**	0.0407	-0.5763	0.9942	-0.5763	0.9942
d_emp_2	0.03234	0.8357	-1.9493	0.9731	-1.9493	0.9731
d_emp_3	0.1122	0.5308	-2.5192	0.9544	-2.5192	0.9544
d_sal_1	-0.02537	0.8609	-0.5057	0.9877	-0.5057	0.9877
d_sal_2	-0.03457	0.8163	0.01897	0.9995	0.01897	0.9995
d_sal_3	0.1102	0.4487	0.05929	0.9988	0.05929	0.9988
d_prod_1	0.01164	0.9377	0.4387	0.9881	0.4387	0.9881
d_prod_2	0.01096	0.9289	-0.3584	0.9909	-0.3584	0.9909
d_prod_3	0.1294	0.3091	-0.04172	0.999	-0.04172	0.999
d_dti_1	0.4573	0.1301	0.4441	0.9974	0.4441	0.9974
d_dti_2	-0.02843	0.8975	-0.2392	0.9985	-0.2392	0.9985
d_dti_3	-0.06039	0.7406	-0.2718	0.9982	-0.2718	0.9982
d_dti_4	0.00325	0.9863	0.3246	0.9979	0.3246	0.9979
d_lti_1	-0.2147	0.4156	0.3089	0.9964	0.3089	0.9964
d_lti_2	0.3753	0.0853	-0.07601	0.999	-0.07601	0.999
d_lti_3	0.1143	0.5509	0.1459	0.9983	0.1459	0.9983
d_lti_4	0.01209	0.9456	-0.07355	0.999	-0.07355	0.999
d_inter_1	0.1995	0.1641	-0.05264	0.9989	-0.05264	0.9989
d_inter_2	0.00796	0.9447	-1.2614	0.9767	-1.2614	0.9767
d_mst_1	0.08535	0.4213	-0.4783	0.988	-0.4783	0.988
d_rel_1	0.1028	0.4215	-1.1393	0.9671	-1.1393	0.9671
d_rel_2	0.1627	0.2597	-0.2643	0.9935	-0.2643	0.9935
d_hst_1	-0.07075	0.4708	-1.7113	0.9474	-1.7113	0.9474
d_edu_1	-0.3305	0.0194	-2.5446	0.9371	-2.5446	0.9371
d_rea_1	-0.1278	0.2936	-1.4545	0.9659	-1.4545	0.9659
d_rea_2	-	0.0165	0.1171	0.9976	0.1171	0.9976
d0	1.6586 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 25. Model Beta T4

Dependent Variable: Recovery Rate						
Beta T4						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-Value	Estimate	P-Value
Intercept	0.536	0.1271	-2.1234	0.9523	-2.1234	0.9523
d_gen_1	0.1706*	0.0787	-0.4283	0.9679	-0.4283	0.9679
d_age_1	0.06384	0.6844	-0.02173	0.9989	-0.02173	0.9989
d_age_2	0.0817	0.5639	0.1036	0.9935	0.1036	0.9935
d_age_3	0.1941	0.2005	0.5197	0.9725	0.5197	0.9725
d_emp_1	0.2953**	0.0148	-0.1611	0.9918	-0.1611	0.9918
d_emp_2	-0.03061	0.8467	-1.1795	0.9458	-1.1795	0.9458
d_emp_3	-0.01067	0.9418	-1.5916	0.9112	-1.5916	0.9112
d_sal_1	-0.03811	0.7986	-0.106	0.9936	-0.106	0.9936
d_sal_2	0.1343	0.3629	0.2767	0.9834	0.2767	0.9834
d_sal_3	0.2	0.2733	0.3105	0.9821	0.3105	0.9821
d_prod_1	0.06119	0.6831	0.5882	0.9639	0.5882	0.9639
d_prod_2	0.01615	0.899	0.01664	0.9989	0.01664	0.9989
d_prod_3	0.1402	0.2905	0.2403	0.985	0.2403	0.985
d_dti_1	0.5279*	0.0858	0.5789	0.9841	0.5789	0.9841
d_dti_2	0.02409	0.9143	0.09656	0.9967	0.09656	0.9967
d_dti_3	-0.04064	0.8272	0.06063	0.9977	0.06063	0.9977
d_dti_4	0.01331	0.9442	0.5141	0.9818	0.5141	0.9818
d_lti_1	-0.251	0.344	0.479	0.9845	0.479	0.9845
d_lti_2	0.2963	0.1777	0.2074	0.9926	0.2074	0.9926
d_lti_3	0.07925	0.6806	0.3818	0.9853	0.3818	0.9853
d_lti_4	0.00316	0.986	0.2134	0.9916	0.2134	0.9916
d_inter_1	0.1451	0.3193	0.2361	0.9853	0.2361	0.9853
d_inter_2	-0.02591	0.8257	-0.6896	0.959	-0.6896	0.959
d_mst_1	0.06637	0.5325	-0.09548	0.9925	-0.09548	0.9925
d_rel_1	0.07527	0.5583	-0.59	0.959	-0.59	0.959
d_rel_2	0.1165	0.4244	0.06863	0.9959	0.06863	0.9959
d_hst_1	-0.07315	0.4673	-1.0103	0.9138	-1.0103	0.9138
d_edu_1	-0.264*	0.0644	-1.6259	0.8869	-1.6259	0.8869
d_rea_1	-0.1316	0.2891	-0.8344	0.9489	-0.8344	0.9489
d_rea_2	-0.3477**	0.0212	0.3523	0.9791	0.3523	0.9791
corr01	-0.01446	0.9111	-1.5314	0.8979	-1.5314	0.8979
corr02	-0.09982	0.5296	-1.6943	0.9104	-1.6943	0.9104
corr021	0.01537	0.8853	0.00067	1	0.00067	1
corr03	0.1595	0.1304	-0.5225	0.9654	-0.5225	0.9654
corr04	0.05341	0.6375	0.183	0.9886	0.183	0.9886
corr05	0.1796	0.826	0.9913	0.989	0.9913	0.989
corr06	0.03428	0.7551	-0.6602	0.9585	-0.6602	0.9585
d_cor7_1	-0.2304	0.2144	0.462	0.9803	0.462	0.9803
d_cor7_2	-0.05913	0.8137	0.8297	0.9645	0.8297	0.9645

Dependent Variable: Recovery Rate						
Beta T4						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-Value	Estimate	P-Value
d_cor7_3	0.06762	0.6935	0.02968	0.9986	0.02968	0.9986
d_cor7_4	-0.1467	0.3915	0.06418	0.9969	0.06418	0.9969
d_cor7_5	-0.1944	0.4286	0.8139	0.9676	0.8139	0.9676
d0	1.6787 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 26. Model Beta T5

Dependent Variable: Recovery Rate						
Beta T5						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-Value	Estimate	P-Value
Intercept	0.443	0.1684	-4.1362	0.998	-4.1362	0.998
d_gen_1	0.152	0.1165	-1.4178	0.9972	-1.4178	0.9972
d_age_1	0.08315	0.5913	-0.6037	0.9993	-0.6037	0.9993
d_age_2	0.09797	0.4827	-0.4381	0.9993	-0.4381	0.9993
d_age_3	0.2139	0.151	0.1678	0.9998	0.1678	0.9998
d_sal_1	0.2654**	0.0265	-1.0088	0.9992	-1.0088	0.9992
d_sal_2	-0.0239	0.8794	-2.7095	0.9973	-2.7095	0.9973
d_sal_3	0.1525	0.4002	-3.3533	0.9959	-3.3533	0.9959
d_emp_1	-0.01748	0.9035	-0.7628	0.9988	-0.7628	0.9988
d_emp_2	-0.01833	0.9015	-0.1888	0.9997	-0.1888	0.9997
d_emp_3	0.1231	0.3996	-0.174	0.9998	-0.174	0.9998
d_prod_1	0.06634	0.6563	0.2931	0.9995	0.2931	0.9995
d_prod_2	0.03373	0.7889	-0.616	0.9991	-0.616	0.9991
d_prod_3	0.1266	0.3325	-0.2598	0.9996	-0.2598	0.9996
d_dti_1	0.4131	0.1738	0.361	0.9998	0.361	0.9998
d_dti_2	-0.00868	0.9688	-0.4684	0.9998	-0.4684	0.9998
d_dti_3	-0.06666	0.7187	-0.5642	0.9997	-0.5642	0.9997
d_dti_4	0.02734	0.8857	0.1997	0.9999	0.1997	0.9999
d_lti_1	-0.215	0.4139	0.2	0.9998	0.2	0.9998
d_lti_2	0.3518	0.1078	-0.2859	0.9997	-0.2859	0.9997
d_lti_3	0.06356	0.7403	-0.00676	1	-0.00676	1
d_lti_4	-0.02256	0.8994	-0.3232	0.9997	-0.3232	0.9997
d_inter_1	0.2111	0.1408	-0.2108	0.9997	-0.2108	0.9997
d_inter_2	0.02115	0.8537	-1.8221	0.9977	-1.8221	0.9977
d_mst_1	0.06092	0.5636	-0.7818	0.9985	-0.7818	0.9985
d_rel_1	0.07554	0.5536	-1.6994	0.9965	-1.6994	0.9965
d_rel_2	0.1112	0.4415	-0.4648	0.9993	-0.4648	0.9993
d_hst_1	-0.04483	0.6537	-2.3454	0.9955	-2.3454	0.9955
d_edu_1	-0.2761**	0.0498	-3.3399	0.9944	-3.3399	0.9944
d_rea_1	-0.07662	0.5292	-2.0756	0.996	-2.0756	0.996
d_rea_2	-0.2881*	0.0542	-0.02927	1	-0.02927	1
scdist01	-0.034	0.4732	0.3938	0.9986	0.3857	0.9991
scdist02	0.03177	0.5196	0.3857	0.9991	0.4874	0.9979
scdist021	0.01717	0.7183	0.4874	0.9979	1.7631	0.9954
scdist03	-0.0916	0.2931	1.7631	0.9954	-0.5934	0.998
scdist04	0.01722	0.7385	-0.5934	0.998	-1.0777	0.9957
scdist05	-0.00597	0.9104	-1.0777	0.9957	0.00075	1
scdist06	-0.04789	0.4106	0.00075	1	0.1092	0.9999
d_cor7_1	-0.164	0.3681	0.1092	0.9999	0.6883	0.9993
d_cor7_2	0.01258	0.9604	0.6883	0.9993	-0.6	0.9991

Dependent Variable: Recovery Rate						
Beta T5						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-Value	Estimate	P-Value
d_cor7_3	0.1202	0.4721	-0.6	0.9991	-0.4894	0.9995
d_cor7_4	-0.09047	0.586	-0.4894	0.9995	0.6813	0.9993
d_cor7_5	-0.1521	0.5316	0.6813	0.9993	0.3938	0.9986
d0	1.6967 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 27. Model Beta T6

Dependent Variable: Recovery Rate						
Beta T6						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-Value	Estimate	P-Value
Intercept	0.3901	0.2462	-4.3821	0.9988	-4.3821	0.9988
d_gen_1	0.1706*	0.078	-1.4722	0.9982	-1.4722	0.9982
d_age_1	0.05321	0.7321	-0.8066	0.9994	-0.8066	0.9994
d_age_2	0.08098	0.5651	-0.5373	0.9994	-0.5373	0.9994
d_age_3	0.227	0.1319	0.1662	0.9999	0.1662	0.9999
d_sal_1	0.2878**	0.0166	-1.0158	0.9996	-1.0158	0.9996
d_sal_2	-0.03304	0.8348	-2.7585	0.9985	-2.7585	0.9985
d_sal_3	0.1912	0.2975	-3.487	0.9973	-3.487	0.9973
d_emp_1	-0.00416	0.977	-0.9184	0.9991	-0.9184	0.9991
d_emp_2	-0.02252	0.8796	-0.251	0.9997	-0.251	0.9997
d_emp_3	0.1353	0.3526	-0.1998	0.9998	-0.1998	0.9998
d_prod_1	0.06142	0.6807	0.2813	0.9996	0.2813	0.9996
d_prod_2	0.02551	0.8387	-0.7382	0.9993	-0.7382	0.9993
d_prod_3	0.1303	0.3225	-0.3254	0.9996	-0.3254	0.9996
d_dti_1	0.4521	0.1387	0.2921	0.9999	0.2921	0.9999
d_dti_2	-0.0161	0.9422	-0.5797	0.9998	-0.5797	0.9998
d_dti_3	-0.08138	0.6578	-0.6195	0.9998	-0.6195	0.9998
d_dti_4	-0.0257	0.8918	0.1387	1	0.1387	1
d_lti_1	-0.2441	0.3573	0.1184	0.9999	0.1184	0.9999
d_lti_2	0.3395	0.1237	-0.3658	0.9998	-0.3658	0.9998
d_lti_3	0.07754	0.6883	-0.09155	1	-0.09155	1
d_lti_4	-0.01043	0.9537	-0.3646	0.9998	-0.3646	0.9998
d_inter_1	0.2012	0.1634	-0.3399	0.9997	-0.3399	0.9997
d_inter_2	0.01139	0.9217	-1.8767	0.9985	-1.8767	0.9985
d_mst_1	0.07787	0.4633	-0.8929	0.9991	-0.8929	0.9991
d_rel_1	0.07531	0.559	-1.7368	0.9979	-1.7368	0.9979
d_rel_2	0.1195	0.4118	-0.6088	0.9994	-0.6088	0.9994
d_hst_1	-0.06362	0.5195	-2.4645	0.9962	-2.4645	0.9962
d_edu_1	-0.2985**	0.037	-3.5276	0.9959	-3.5276	0.9959
d_rea_1	-0.1128	0.3535	-2.131	0.9978	-2.131	0.9978
d_rea_2	-0.337**	0.0234	-0.1263	0.9999	-0.1263	0.9999
score	0.4047*	0.0559	-1.1732	0.9992	-1.1732	0.9992
d_cor7_1	-0.2217	0.2222	0.05549	1	0.05549	1
d_cor7_2	-0.02503	0.9199	0.7007	0.9996	0.7007	0.9996
d_cor7_3	0.06944	0.6749	-0.6769	0.9994	-0.6769	0.9994
d_cor7_4	-0.1423	0.3911	-0.5918	0.9997	-0.5918	0.9997
d_cor7_5	-0.1937	0.422	0.6719	0.9996	0.6719	0.9996
d0	1.6698 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 28. Model Beta T7

Dependent Variable: Recovery Rate						
Beta T7						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
Intercept	0.2102	0.7669	-2.0937	0.9989	-2.0937	0.9989
d_gen_1	0.1642*	0.086	-0.4396	0.9988	-0.4396	0.9988
d_age_1	0.118	0.4325	-0.01429	1	-0.01429	1
d_age_2	0.07576	0.579	0.1304	0.9998	0.1304	0.9998
d_age_3	0.2121	0.1508	0.5315	0.9989	0.5315	0.9989
d_sal_1	0.2244*	0.0591	-0.184	0.9996	-0.184	0.9996
d_sal_2	0.02357	0.8799	-1.2324	0.998	-1.2324	0.998
d_sal_3	0.1352	0.4483	-1.6335	0.9971	-1.6335	0.9971
d_emp_1	-0.04071	0.7773	-0.09027	0.9999	-0.09027	0.9999
d_emp_2	-0.03284	0.8239	0.2922	0.9994	0.2922	0.9994
d_emp_3	0.1107	0.4437	0.3186	0.9996	0.3186	0.9996
d_prod_1	-0.00667	0.9641	0.6076	0.9988	0.6076	0.9988
d_prod_2	-0.00199	0.987	0.01946	1	0.01946	1
d_prod_3	0.1159	0.3628	0.2505	0.9996	0.2505	0.9996
d_dti_1	0.4355	0.156	0.6329	0.9994	0.6329	0.9994
d_dti_2	-0.01779	0.9362	0.1047	0.9999	0.1047	0.9999
d_dti_3	-0.07982	0.666	0.04606	0.9999	0.04606	0.9999
d_dti_4	0.0216	0.9098	0.5145	0.9992	0.5145	0.9992
d_lti_1	-0.1932	0.4712	0.5307	0.9995	0.5307	0.9995
d_lti_2	0.3502	0.1104	0.207	0.9998	0.207	0.9998
d_lti_3	0.0696	0.721	0.3867	0.9995	0.3867	0.9995
d_lti_4	0.01706	0.9237	0.1994	0.9997	0.1994	0.9997
d_inter_1	0.223	0.1204	0.2842	0.9995	0.2842	0.9995
d_inter_2	0.04637	0.6875	-0.7051	0.9986	-0.7051	0.9986
d_mst_1	0.06475	0.5367	-0.06561	0.9998	-0.06561	0.9998
d_rel_1	0.08686	0.5038	-0.6183	0.9988	-0.6183	0.9988
d_rel_2	0.1637	0.2602	0.1139	0.9998	0.1139	0.9998
d_hst_1	-0.05582	0.5667	-1.0282	0.9976	-1.0282	0.9976
d_edu_1	-0.3188**	0.0232	-1.6155	0.9952	-1.6155	0.9952
d_rea_1	-0.1121	0.3581	-0.858	0.9983	-0.858	0.9983
d_rea_2	-0.2952*	0.0515	0.4123	0.9993	0.4123	0.9993
dq_12_1	0.2068	0.4958	-1.6659	0.9976	-1.6659	0.9976
dq_12_2	0.2275	0.5158	0.7005	0.9991	0.7005	0.9991
dq_13_1	0.3833	0.2332	-1.8788	0.9972	-1.8788	0.9972
dq_13_2	-0.09736	0.8338	0.9049	0.9989	0.9049	0.9989
dq_14_1	0.08008	0.5111	0.1398	0.9998	0.1398	0.9998
dq_14_2	0.00355	0.9785	0.4535	0.9994	0.4535	0.9994
dq_15_1	0.0352	0.9204	-1.9555	0.9965	-1.9555	0.9965
dq_15_2	-0.3059	0.5495	0.9476	0.9991	0.9476	0.9991
dq_16_1	0.1465	0.389	-0.8312	0.9989	-0.8312	0.9989

Dependent Variable: Recovery Rate						
Beta T7						
	Error		Zero		One	
Variables	Estimate	P-	Estimate	P-	Estimate	P-
dq_16_2	0.229	0.2083	0.1126	0.9998	0.1126	0.9998
dq_17_1	-0.1588	0.4697	-1.7425	0.9963	-1.7425	0.9963
dq_17_2	0.00063	0.9986	0.8964	0.9987	0.8964	0.9987
dq_18_1	-0.05901	0.9156	-2.0681	0.9987	-2.0681	0.9987
dq_18_2	-0.06378	0.9528	0.9915	0.9995	0.9915	0.9995
dq_19_1	0.2019	0.3311	0.2675	0.9996	0.2675	0.9996
dq_19_2	0.08511	0.6499	-1.1653	0.9981	-1.1653	0.9981
dq_20_1		0.0541	-1.4287	0.9981	-1.4287	0.9981
dq_20_2	-	0.0057	0.5909	0.9992	0.5909	0.9992
dq_21_1	0.1994	0.3135	-2.0597	0.8112	-2.0597	0.8112
dq_21_2	-0.3949	0.1225	0.8809	0.9593	0.8809	0.9593
d0	1.7202 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 29. Model Beta T8

Dependent Variable: Recovery Rate						
Beta T8						
	Error		Zero		One	
Variables	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Intercept	0.2061	0.7868	-2.1016	0.9995	-2.1016	0.9996
d_gen_1	0.1608	0.1039	-1.317	0.9755	-0.4403	0.9994
d_age_1	0.1061	0.5017	-0.01546	1	-0.01556	1
d_age_2	0.1075	0.4467	0.1256	0.9999	0.1256	0.9999
d_age_3	0.2967*	0.0503	0.5299	0.9994	0.53	0.9995
d_sal_1	0.2616**	0.0314	-0.1856	0.9998	-0.1855	0.9998
d_sal_2	-0.08731	0.5879	-1.2352	0.9987	-1.2352	0.9989
d_sal_3	0.2408	0.1951	-1.6407	0.9985	-1.6407	0.9987
d_emp_1	-0.06806	0.6422	-0.09227	0.9999	-0.09237	0.9999
d_emp_2	-0.04662	0.7566	0.2864	0.9997	0.2864	0.9997
d_emp_3	0.1227	0.4116	0.3183	0.9997	0.3184	0.9997
d_prod_1	0.04745	0.7569	0.6055	0.9994	0.6055	0.9994
d_prod_2	-0.04235	0.7424	0.01471	1	0.01467	1
d_prod_3	0.09852	0.4587	0.2502	0.9998	0.2502	0.9998
d_dti_1	0.4889	0.1227	0.6307	0.9997	0.6307	0.9998
d_dti_2	-0.00394	0.9863	0.102	0.9999	0.102	0.9999
d_dti_3	-0.1212	0.5293	0.04571	1	0.04559	1
d_dti_4	0.00736	0.97	0.5134	0.9996	0.5135	0.9996
d_lti_1	-0.3048	0.2645	0.528	0.9998	0.528	0.9998
d_lti_2	0.2947	0.1905	0.2047	0.9999	0.2046	0.9999
d_lti_3	0.0078	0.9689	0.3865	0.9997	0.3865	0.9997
d_lti_4	0.02187	0.905	0.1971	0.9998	0.1971	0.9999
d_inter_1	0.1949	0.1802	0.2803	0.9997	0.2804	0.9997
d_inter_2	0.04761	0.6871	-0.7085	0.9991	-0.7086	0.9992
d_mst_1	0.00615	0.954	-0.06919	0.9999	-0.06928	0.9999
d_rel_1	0.07705	0.5585	-0.6193	0.9994	-0.6193	0.9994
d_rel_2	0.1458	0.3232	0.1105	0.9999	0.1104	0.9999
d_hst_1	-0.04899	0.6305	-1.0345	0.9984	-1.0345	0.9985
d_edu_1	-0.2702*	0.0579	-1.6218	0.9978	-1.6219	0.9981
d_rea_1	-0.107	0.3919	-0.8654	0.999	-0.8654	0.9991
d_rea_2	-0.2748*	0.0791	0.4114	0.9996	0.4114	0.9997
scdist01	0.00718	0.8838	0.9129	0.9974	0.9126	0.9977
scdist02	-0.00893	0.859	0.9165	0.9968	0.9161	0.9972
scdist021	0.01607	0.743	0.9632	0.9978	0.9631	0.9982
scdist03	-0.1069	0.2351	1.6074	0.9971	1.6077	0.9975
scdist04	0.05681	0.2829	0.4172	0.9991	0.4167	0.9993
scdist05	-0.00167	0.9753	0.1047	0.9998	0.1047	0.9998
scdist06	-0.09836	0.1115	0.868	0.9982	0.8678	0.9984
d_cor7_1	-0.2624	0.185	0.4606	0.9997	0.4606	0.9997

Dependent Variable: Recovery Rate						
Beta T8						
	Error		Zero		One	
Variables	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
d_cor7_2	0.01142	0.9654	0.8379	0.9994	0.8379	0.9995
d_cor7_3	0.056	0.7541	0.01667	1	0.0167	1
d_cor7_4	-0.1618	0.355	0.09473	0.9999	0.09469	0.9999
d_cor7_5	-0.298	0.2489	0.8208	0.9995	0.8208	0.9996
dq_12_1	0.2285	0.4553	-1.6679	0.9988	-1.6679	0.9989
dq_12_2	0.2602	0.4623	0.6978	0.9996	0.6979	0.9997
dq_13_1	0.3354	0.3155	-1.8817	0.9979	-1.8818	0.9982
dq_13_2	0.0148	0.9885	0.9032	0.9999	0.9032	0.9999
dq_14_1	0.04727	0.707	0.1394	0.9999	0.1395	0.9999
dq_14_2	-0.00297	0.9823	0.4516	0.9996	0.4516	0.9997
dq_15_1	0.01007	0.9775	-1.9585	0.9988	-1.9585	0.9989
dq_15_2	-0.2891	0.5729	0.946	0.9996	0.946	0.9996
dq_16_1	0.1417	0.4105	-0.832	0.9993	-0.8319	0.9994
dq_16_2	0.2622	0.1579	0.1072	0.9999	0.1071	0.9999
dq_17_1	-0.1564	0.4811	-1.7441	0.9987	-1.7441	0.9989
dq_17_2	0.01517	0.9663	0.8938	0.9994	0.8938	0.9994
dq_18_1	-0.1355	0.8116	-2.0759	0.9992	-2.0759	0.9993
dq_18_2	-0.07315	0.9478	0.9914	0.9997	0.9914	0.9997
dq_19_1	0.3166	0.1401	0.2671	0.9998	0.2672	0.9999
dq_19_2	0.1833	0.3333	-1.1726	0.999	-1.1727	0.9991
dq_20_1	-0.5314**	0.0124	-1.4312	0.9991	-1.4311	0.9992
dq_20_2	-0.7688***	0.0014	0.5872	0.9996	0.5872	0.9996
dq_21_1	0.2445	0.2511	-1.9276	0.9987	-1.9276	0.9988
dq_21_2	-0.232	0.8	0.8947	0.9999	0.8947	0.9999
d0	1.7376 <.0001					
***, **, * indicate 1% ,5% and 10% significance levels, respectively.						

Appendix 30. Recovery Rate Models Overdue Balance

Variables	Model 1 Overdue Balance		Model 2 Overdue Balance		Exploratory Variables	Model 1 Overdue Balance		Model 2 Overdue Balance	
	coefficient	P-Value	coefficient	P-Value		coefficient	P-Value	coefficient	P-Value
Constant	-463.6294	0.000	-389.3774	0.000	l1_call1	-256.1588	0.000	-359.5600	0.000
d_gen_1	0.9121282	0.071	-1.0993	0.132	l1_sms	-260.9699	0.000	-367.5741	0.000
d_age_1	41.64924	0.000	159.8126	0.000	l1_spc	-57.53807	0.000	-108.8432	0.000
d_age_2	4297944	0.000	166.1115	0.000	l1_call2	-197.3801	0.000	-316.1666	0.000
d_age_3	25.7202	0.000	94.7893	0.000	l1_col1	-159.551	0.000	-209.9550	0.000
d_age_4	13.63803	0.000	55.1911	0.000	l1_col2	-105.4304	0.000	-121.4622	0.000
d_sal_1	6.512973	0.004	-168.6496	0.000	l1_col3	56.05598	0.000	79.3606	0.000
d_sal_2	6.078524	0.006	-163.3286	0.000	l2_sms	-185.8208	0.000	-285.1386	0.000
d_sal_3	1372961	0.000	-94.9097	0.000	l2_call2	-179.2883	0.000	-291.5136	0.000
d_sal_4	7.248032	0.002	-87.1616	0.000	l2_col1	-164.6274	0.000	-212.8397	0.000
d_sal_5	11.77858	0.000	-65.1471	0.000	l2_col2	-110.4063	0.000	-124.8195	0.000
d_sal_6	8.210791	0.006	-11.1957	0.000	l2_col3	52.48934	0.000	77.7802	0.000
d_inter_1	-43.95138	0.000	-109.2726	0.000	l3_call2	-136.9423	0.000	-229.5463	0.000
d_inter_2	-23.9792	0.000	-49.1204	0.000	l3_col1	-165.6753	0.000	-211.5525	0.000
d_inter_3	-12.34821	0.000	-34.9173	0.000	l3_col2	-113.7493	0.000	-127.2828	0.000
d_emp_1	25.18909	0.000	79.3555	0.000	l3_col3	48.45431	0.000	75.6040	0.000
d_emp_2	14.09373	0.000	48.2902	0.000	l4_col1	-113.1542	0.000	-157.3697	0.000
d_emp_3	8.451349	0.000	49.8974	0.000	l4_col2	-62.76525	0.000	-76.7450	0.000
d_emp_4	1.620075	0.472	27.1767	0.000	l4_col3	89.29336	0.000	117.5694	0.000
d_emp_5	-8.122247	0.001	24.2855	0.000	l5_col2	-40.14955	0.000	-53.0336	0.000
d_prod_1	13.46657	0.000	57.0732	0.000	l5_col3	102.8579	0.000	133.1309	0.000
d_prod_2	-7.547975	0.000	-23.8786	0.000	l6_col2	-42.72701	0.000	-53.9376	0.000
d_prod_3	15.02462	0.000	49.4895	0.000	l6_col3	99.08024	0.000	130.1663	0.000
d_dti_1	-157.733	0.000	-132.6172	0.000	l7_col2	-45.69565	0.000	-54.3547	0.000
d_dti_2	-133.2521	0.000	-89.7592	0.000	l7_col3	95.66453	0.000	128.7635	0.000
d_dti_3	-85.63341	0.000	-46.5119	0.000	l8_col2	-23.03841	0.000	-31.4952	0.000
d_dti_4	- 41.378561	0.000	-2.3748	0.000	l8_col3	115.416	0.000	148.9223	0.000
d_lti_1	228.8753	0.000	95.4013	0.000	l9_col2	15.42757	0.000	7.0828	0.000
d_lti_2	185.0189	0.000	79.8201	0.000	l9_col3	144.1119	0.000	178.7295	0.000
d_lti_3	152.0052	0.000	67.6280	0.000	l10_col2	12.58423	0.000	7.4236	0.000
d_lti_4	109.2793	0.000	57.0238	0.000	l10_col3	140.5182	0.000	175.8913	0.000
d_mst_1	45.10767	0.000	149.3763	0.000	l11_col2	9.126362	0.000	4.4572	0.000
d_landline	5.608163	0.411	16.1041	0.000	l11_col3	136.7952	0.000	172.9285	0.000
d_mobile	-89.65093	0.000	-162.8257	0.000	l12_col2	13.61649	0.000	10.2389	0.044
Balance	0.7517179	0.000			l12_col3	140.4618	0.000	177.5205	0.000
loan_amount			0.4728	0.000	Week	3.065381	0.000	2.8128	0.000

Appendix 31. Recovery Rate Models Rec Ratio and Bal Ratio

Variables	Model 3		Model 4		Exploratory Variables	Model 3		Model 4	
	Rec Ratio		Bal Ratio			Rec Ratio		Bal Ratio	
	coefficient	P-Value	coefficient	P-Value		coefficient	P-Value	coefficient	P-Value
Constant	0.2243901	0.000	0.0761501	0.028	l1_call1	-0.2352919	0.000	-0.1855145	0.000
d_gen_1	-0.0088598	0.000	0.0559932	0.000	l1_sms	-0.2378711	0.000	-0.1857144	0.000
d_age_1	0.1185028	0.000	0.0812587	0.000	l1_spc	-0.0560485	0.000	0.0133532	0.305
d_age_2	0.1110319	0.000	0.1345238	0.000	l1_call2	-0.1788002	0.000	-0.0301284	0.003
d_age_3	0.0623442	0.000	0.0512079	0.000	l1_col1	-0.093368	0.000	0.0913107	0.000
d_age_4	0.0421257	0.000	0.0225932	0.000	l1_col2	-0.0281891	0.000	0.1562791	0.000
d_sal_1	-0.0594774	0.000	-0.0440380	0.000	l1_col3	0.1123439	0.000	0.3024755	0.000
d_sal_2	-0.0704149	0.000	-0.0383450	0.000	l2_sms	-1.172199	0.000	-0.0924785	0.000
d_sal_3	-0.0398707	0.000	0.1064005	0.000	l2_call2	-0.1571771	0.000	0.0091741	0.389
d_sal_4	-0.0260299	0.000	-0.0092028	0.411	l2_col1	-0.0942613	0.000	0.0912894	0.000
d_sal_5	-0.0257161	0.000	-0.0108287	0.360	l2_col2	-0.0299007	0.000	0.1539570	0.000
d_sal_6	0.0035829	0.097	0.1970246	0.000	l2_col3	0.1111445	0.000	0.2983007	0.000
d_inter_1	-0.0656796	0.000	-0.0712231	0.000	l3_call2	-0.1131232	0.000	0.0441174	0.000
d_inter_2	-0.0379341	0.000	-0.0576838	0.000	l3_col1	-0.0928734	0.000	0.0866164	0.000
d_inter_3	-0.0312297	0.000	-0.0511319	0.000	l3_col2	-0.0313716	0.000	0.1489109	0.000
d_emp_1	0.0394107	0.000	0.0321005	0.000	l3_col3	0.1102686	0.000	0.2947753	0.000
d_emp_2	0.0122767	0.000	-0.0154688	0.130	l4_col1	-0.0513149	0.000	0.1456111	0.000
d_emp_3	0.0186772	0.000	0.0165174	0.071	l4_col2	0.0054138	0.000	0.1990546	0.000
d_emp_4	0.0053515	0.001	-0.0162424	0.127	l4_col3	0.1358313	0.000	0.3226457	0.000
d_emp_5	0.0219847	0.000	-0.0073665	0.511	l5_col2	0.021274	0.000	0.2116013	0.000
d_prod_1	0.0452778	0.000	-0.0308003	0.000	l5_col3	0.1448559	0.000	0.3311655	0.000
d_prod_2	0.0022047	0.000	-0.0551504	0.000	l6_col2	0.0205486	0.000	0.2052152	0.000
d_prod_3	0.0195924	0.000	-0.0355601	0.000	l6_col3	0.1434092	0.000	0.3268552	0.000
d_dti_1	-0.1407517	0.000	-0.2875122	0.000	l7_col2	0.0196409	0.000	0.2015645	0.000
d_dti_2	-0.0722938	0.000	-0.2176624	0.000	l7_col3	0.1422687	0.000	0.3239811	0.000
d_dti_3	-0.0282858	0.000	-0.1338421	0.000	l8_col2	0.0359277	0.000	0.2214927	0.000
d_dti_4	-0.0042823	0.000	0.0470426	0.000	l8_col3	0.1535894	0.000	0.3344245	0.000
d_lti_1	0.1207919	0.000	0.3169189	0.000	l9_col2	0.0627189	0.000	0.2575298	0.000
d_lti_2	0.083361	0.000	0.2356927	0.000	l9_col3	0.1704406	0.000	0.3522831	0.000
d_lti_3	0.0581456	0.000	0.2603729	0.000	l10_col2	0.0627536	0.000	0.2516351	0.000
d_lti_4	0.0383701	0.000	0.0755739	0.000	l10_col3	0.1686618	0.000	0.3479532	0.000
d_mst_1	0.0682025	0.000	0.0614979	0.000	l11_col2	0.061636	0.000	0.2485351	0.000
d_landline	-0.0189771	0.000	0.0374457	0.246	l11_col3	0.1667408	0.000	0.3435245	0.000
d_mobile	-0.0965545	0.000	-0.0733839	0.000	l12_col2	0.0649245	0.000	0.2483073	0.000
					l12_col3	0.1690164	0.000	0.3442698	0.000
					Week	0.001854	0.000	0.0037975	0.000