Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

A battle of pooled and panel data in credit risk modelling

Thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements
for the degree

MASTER OF SCIENCE
IN
APPLIED MATHEMATICS

by

Gavriella Michael
4627466

Delft, the Netherlands
Wednesday 23rd January, 2019

MSc thesis APPLIED MATHEMATICS
CONFIDENTIAL

"A battle of pooled and panel data in credit risk modelling"

Gavriella Michael

Delft, the Netherlands

Daily supervisor

Dr. K. Kuoch

University supervisor

Dr. P. Cirillo

Other thesis committee members

Prof.dr.ir. C.W. Oosterlee

Wednesday 23rd January, 2019 Delft, the Netherlands

Abstract

When dealing with datasets where the observations are obtained from the same cross-sectional units at multiple time points, most of the times, heterogeneity arises across he cross-sectional units. If one ignores this heterogeneity, assuming that the data are pooled, the parameters estimations run the risk of being inconsistent. This thesis studies the difference between panel data and pooled data models with regard to their construction procedure and their predictive performance.

An application is discussed per credit risk modelling for a mortgage portfolio. Therein, different models were constructed, covering pooled and panel linear models and pooled and panel logistic models. By model performance and testing comparison, we found that by adding the heterogeneity effect in the regression model the discriminatory power is improved. At the same time, however, it provides lower predicted losses than the observed ones. We have also noted that, most of the times, the pooled model fails to estimate accurate predictions.

This thesis has been carried out jointly with TU Delft / Department of Applied Mathematics and the Central Risk Management / Model Validation department of ABN AMRO Bank.

Acknowledgements

I would first like to express my sincere gratitude to my supervisors Dr. Pasquale Cirillo and Dr. Kevin Kuoch for the continuous support of my Master thesis research, for his patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisors and mentors for my Mater thesis study.

Besides my advisors, I wish to thank the member of my dissertation committee Prof.Dr.ir. C.W. Oosterlee for generously offering his time, support, guidance and good will throughout the preparation and review of this document.

I must express my very profound appreciation to my parents and to my friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this Master thesis. This accomplishment would not have been possible without them. Thank you.

Finally, I would like to thank my friend Kyriakos Demetriou for being so supportive while I was working on my thesis. Thanks for your emotional support for being so understanding and for being such a great friend.

Contents

1

Introduction

The Basel Committee on Banking Supervision (BCBS) has introduced the 1988 Accords, also known as Basel I. That is a set of recommendations for regulations in the banking industry, focused on financial stability, developing capital requirements for banks based on the riskiness of their financial positions. Since then, the Basel Committee has published several proposals in order to revise the Basel I framework. In 2004, a first revision, Basel II [BCBS, 2006], has introduced the foundation of three pillars: minimum capital requirements, supervisory review and market discipline – aiming to promote a stronger risk management framework.

Financial institutions have placed significant reliance on quantitative analyses and mathematical models in order to assist decision-making (strategy, budgeting, planning, balance sheet steering, etc.). The increasing complexity of quantitative models has given rise to a new type of risk : model risk. Managing model risk has subsequently come under scrutiny from regulators. Basel II [BCBS, 2006] has introduced the requirements of validating models, see [BCBS, 2006, Art.302–305].

The supervision of model risk relies on an independent Model Validation function. Model Validation (thereafter MV), under regulatory guidance, is in charge of monitoring all phases of model development and implementation with the purpose of mitigating model risk. In addition, MV assesses the compliance of models to internal policies and external regulations. Such MV function is yet present in many industries (IT, pharmaceuticals, etc.) although, with banking capital regulation, it tends to be in a more mature stage when it comes to financial institutions than other sectors.

In ABN AMRO Bank, the MV department is in charge of mitigating model risk. It covers numerous model risk dimensions such as data, methodology, implementation and use. The outcome of the validation process affects every level of the organisation – from individual client acceptance to strategic decision making and steering. Despite a broad risk and model landscape, this thesis focuses on a major risk run by financial institutions: credit risk.

One of the components of credit risk is the risk of default on a debt that may arise from a borrower who fails to make the required payments. Bank-issued credit makes up the largest proportion of credit in existence. Thus, banking is about credit creation promoting credit risk as one of the core risks that financial institutions face.

Within Basel II framework, banks can opt for different approaches to assess their credit risk: a Standardised approach, Foundation Internal Rating-Based (IRB) approach and the Advanced Internal Rating-Based approach. Within Standardised Approach, banks have to divide their credit exposures into classes, based on certain observable characteristics of the exposures (e.g. a corporate or mortgage loan). For all classes, a fixed risk weight is determined by the supervisor. The minimum ratio of capital to the total weighted exposure is 8%. Under IRB approaches, four inputs are needed for credit risk determination and capital calculations: the probability of default, the loss given default, the exposure at default and the remaining maturity of the loan. IRB approaches permits a bank to use internal ratings as primary inputs to capital calculations. More precisely, the Foundation IRB Approach allows banks to determine the probability of default for each borrower whilst the supervisor supplies the other inputs (loss given default, the exposure at default and the maturity). On the other hand, the Advanced IRB Approach permits banks to estimate all four inputs needed for credit risk determination and capital calculations, that is, probability of default (PD), loss given default (LGD), the exposure at default (EAD) and the maturity.

Since ABN AMRO Bank uses the Advanced IRB approach to calculate its regulatory capital for credit risk, it is allowed to estimate its own credit risk components and construct its own models. This thesis focuses on the construction of the Loss Given Default model for mortgage loans. The LGD represents the ratio of the Exposure at Default (EAD) expected to be lost if a counterparty goes into default. It is of crucial importance that the LGD estimations are accurate, and the models have adequate discriminatory power. That is because underestimation of the LGD estimations means that extreme losses on the loan portfolio are not covered and, at the same time, overestimation of losses lead the bank to hold additional capital, which does not yield a return. Therefore, it is desirable that both aforementioned situations are avoided and, also, that the LGD model is able to discriminate between high and low losses.

The current LGD models of the bank are based on the standard statistical approaches, like pooled OLS regression and pooled logistic analyses. However, these methods ignore the fact that the LGD dataset for the defaulted contracts has a panel data form. Panel data or longitudinal data are a combination of cross-sectional and time-series data where the observations are obtained from the same cross-sectional units of households, individuals, firms, countries, etc. at multiple points in time. Hence, the datasets containing panel data are two-dimensional with i denoting the cross-section dimension and t the time-series dimension. The LGD datasets of the Bank for the defaulted contracts is consisted of a number of contracts that are observed more than one time. That is because the defaulted contracts belong to the defaulted dataset until they will cure and go to the performing portfolio. Or, in the case of “non-cure” , the debt is considered a loss (write off) or the contract is recovered without loss for the bank (paid-off), and then the contract is moved out from the defaulted portfolio. This process may take months or years. Therefore, these dataset can be considered as panel datasets.

Collecting information from the same individuals or firms over time leads to the natural assumption that these individuals are heterogeneous. The methods of panel data estimation capture and control for this heterogeneity by taking into consideration all the unobserved individual- and time-specific variables. However, the pool data regression models, that the Bank uses, are not able to capture this individual heterogeneity and instead they choose to ignore it. Some people believe that the ignorance is bliss. In this case, this ignorance might lead to biased and inconsistent statistical inferences due to the omitted variables phenomenon. The purpose of the present thesis is therefore to construct the pooled and panel defaulted LGD models and compare their perfomances.

We expect that by adding the unobserved heterogeneity term in the regression model, leading to a panel data model, will immediately improve its accuracy and give predictions closer to the realized ones. Moreover, the ability to discriminate among the low and high loses might me better for the panel data model rather than the pool model.

We formulate the following research question:

- Is the panel data LGD model, for the defaulted contracts, preferable than the pooled data LGD model?

To answer this question we take into account the following sub questions:

- Does the panel data LGD model has better discriminatory power than the pooled data LGD model?

- Does the panel data LGD model gives more accurate predictions than the pooled data LGD model's predictions?

This paper is structured as follows. Chapter 2 describes the mathematical background of the panel data models and explains the methodology that will be used in the next chapter. In Chapter 3, the LGD pooled and panel data models for the defaulted contracts are constructed, and their performance is compared by means of discriminatory and calibration ability. Finally, in Chapter 4, we provide a summary of our conclusions and suggest some ideas for future research.

2

Panel data modelling

Contents

The analysis of datasets that combine cross-sectional and time-series data is one of the most active areas of research in econometrics. These data are called panel or longitudinal, and they contain a number of observations where the same cross-sectional units have been repeatedly observed over different time periods [Verbeek, 2004].

Panel data forms a special case of the so-called pooled data, where each cross-section observation is not necessarily collected from the same unit [Greene, 2012]. An example of panel data is the Panel Study of Income Dynamics (PSID), a collection of the Institute for Social Research at University of Michigan. The first wave, in 1968, interviewed 4800 families while in 2001 the PSID included more than 7000 families. For the first 29 years, from 1968 to 1996, families were interviewed once a year, whereas from 1997 onward data are collected biennially. In 2002, Lundberg and Rose used the panel data from the PSID with regard to the years 1968-1992, in order to study the effects on fatherhood and, consequently, the differential effects on sons and daughters according to the father's labour supply and hourly wages.

The aforementioned economists have obtained inferences by first assuming that the data are pooled data and then panel data, and, based on this, they observed different estimates. These differences may be due to the ability of the panel data models to capture the heterogeneity and the unobservable effects across the different cross sections or time series units [Baltagi, 2013, Chap2]. In addition, panels allow a researcher to study the dynamics of changes and relate the behaviour of a cross-section unit across different time periods, which is very difficult with cross-sectional evidence – see [Baltagi, 2013, Chap1]. For example, panels allow us to estimate the proportion of low paid workers in a population and examine whether this status is transitory or long-lived over the employee's life cycle in the labour market. Cross-sectional data can estimate what proportion of the population's workers is low paid at a certain point in time. Only panel data can show how this proportion changes over time and if a worker who is low paid in a certain period of time will remain low paid in a different one.

The panel data can be categorized as balanced and unbalanced, depending on the number of times that each cross-section unit is conserved or not. If the dataset consists of N different cross-sectional units, for instance N individuals, and each individual is observed throughout all T time periods, then, the dataset is a balanced panel [Greene, 2012]. Therefore, the total number of observations in the panel is NT. However, when some individuals are not observed over all T points in time, but only Ti times, then, the total number of observations in the dataset is $\sum_{i=1}^{N} T_i$, and the dataset is considered to be unbalanced. Moreover, if the panel data regression model manages to capture the unobservable cross-section-specific or time-specific effect among the observations, then, it is called "one-way error component" or "static", whereas, when the model measures both individual and time heterogeneity, it is called "two-way error component" or "dynamic" [Baltagi, 2013, Chap3] [Verbeek, 2004]. This thesis focuses on one-way error component models with cross-sectional specific effect on unbalanced panel data, since this will be the case in the application example in Chapter 3.

In this chapter, different panel data regression models are explained, and tests regarding the most appropriate model for the data will be described. With this in mind, Section 2.1 describes linear panel models, and Section 2.2 considers specific features of binary choice panel models.

2.1 Linear models

A panel data regression model indexes all variables by an i and a t, where i denotes the cross-section dimension (individuals, clients, firms, etc.) and t represents the time-series dimension. The general framework of a one-way error unbalanced panel data model is:

$$y_{it} = \alpha + x'_{it}\beta + u_{it}, \qquad i = 1, ..., N;\, t = 1, ..., Ti, \tag{2.1.1}$$

where y_{it} is the dependent variable, α is a scalar, β is $K \times 1$ and X_{it} is the i-th observation on K explanatory variables. The term u_{it} is the disturbance that consists of the *unobservable* firm-specific effect μ_i and the remainder disturbance v_{it}, i.e.

$$u_{it} = \mu_i + v_{it}. \tag{2.1.2}$$

It is important to note that μ_i does not vary over time, and it captures any unit-specific effect which is not included in the regression, whereas the term v_{it} does vary over time and unit; hence, it can be considered as the usual disturbance in the regression.

To estimate the above model, it is essential to determine the nature of the unobserved variable μ_i. If we treat μ_i as N fixed unknown parameters, the model in 2.1.1 is referred to as the fixed effects model. An alternative approach is treating the μ_i as random and, therefore, transform the model in 2.1.1 into a random effects model. However, if such unobservable effects do not exist in the data, then, the preferable model is the pooled model.

2.1.1 Pooled data

When the individual heterogeneity μ_i does not exist and the model has the following form

$$y_{it} = \alpha + x'_{it}\beta + \epsilon_{it}, \qquad i = 1, ..., N;\ t = 1, ..., Ti, \tag{2.1.3}$$

with strictly exogenous regressors x_{it}, same finite variance σ^2 for all disturbance terms ϵ_{it} (homoscedasticity), uncorrelation among the different error terms (nonautocorrelation), independence across observations i and no multicollinearity among the independent variables, then, the ordinary least squares (thereafter, OLS) estimation method produces consistent and efficient parameter estimators [Greene, 2012, Chapter 10].

The OLS estimator for β, which now includes the constant term α, can derive from

$$b := (\sum_{i,t} x_{it}x'_{it})^{-1} \sum_{i,t} x_{it}y_{it}, \tag{2.1.4}$$

which can be written in a vector form as

$$b := (X'X)^{-1}X'y, \tag{2.1.5}$$

with variance

$$\mathbb{V}(b|X) = \sigma^2(X'X)^{-1}. \tag{2.1.6}$$

2.1.2 Fixed effect model

The fixed effects model is given by

$$y_{it} = \alpha + \mu_i + x'_{it}\beta + v_{it}, \qquad i = 1, ..., N;\ t = 1, ..., Ti, \tag{2.1.7}$$

where we assume that the individual specific effect μ_i is a fixed parameter, the error terms v_{it} are IID(0,σ_v^2) and the x'_{it} is independent of the v_{it} for all i and t. This model is appropriate when we are interested in the behavior of N specific individuals or firms and not on a randomly selected set of N individuals from a large population. This will be discussed later.

The meaning of the term "fixed effects" refers to the fact that even though the fixed variable μ_i varies between units, it remains constant over time. Assuming that the μ_i differs for each entity, we can take into account the uniqueness and peculiarity of each individual. If we consider the fixed effects μ_i as part of the intercept, then, any correlation among them and the regressors is allowed. Therefore, by meeting all OLS assumptions, we can perform least squares dummy variable (LSDV) regression on 2.1.7 or within effect estimation method.

The LSDV estimator is obtained by applying OLS on the fixed effect model, including a dummy variable for each individual i in the model. That is,

$$y_{it} = \alpha + \sum_{j=1}^{N} \mu_i \delta_{ij} + x'_{it}\beta + v_{it}, \tag{2.1.8}$$

where the δ_{ij} is the Kronecker delta. Then, we can obtain estimates of α, β and μ_i. However, a large number of individuals in panel data results in too many individual dummies, and hence complex computations [Verbeek, 2004, Chapter 10].

Fortunately, another strategy can be used in order to obtain the same estimator for β in a simpler way: the within effect estimation. This strategy requires first the calculation of the individual means over time of the dependent and independent variables on 2.1.7 to get

$$\bar{y}_{i.} = \alpha + \mu_i + \beta\bar{x}_{i.} + \bar{v}_{i.}, \tag{2.1.9}$$

and then transform the model in deviations from individual means and perform OLS on the transformed model:

$$y_{it} - \bar{y}_{i.} = \beta(x_{it} - \bar{x}_{i.}) + (v_{it} - \bar{v}_{i.}), \tag{2.1.10}$$

where $\bar{y}_{i.} = \sum_{t=1}^{T_i} y_{it}/T_i$, $\bar{x}_{i.} = \sum_{t=1}^{T_i} x_{it}/T_i$ and $\bar{v}_{i.} = \sum_{t=1}^{T_i} v_{it}/T_i$.

This transformation is called the within transformation, while the resulting regression model 2.1.10 does not include the FE estimators of the individual effects μ_i and the constant term α. The resulting within estimator for β is

$$\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}, \tag{2.1.11}$$

with $Var(\tilde{\beta}) = \sigma_v^2(\tilde{X}'\tilde{X})^{-1}$ where $\tilde{x'_{it}} = x_{it} - \bar{x}_{i.}$ and $\tilde{y} = y_{it} - \bar{y}_{i.}$. Therefore,

$$\tilde{\beta} = (\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_{i.})(x_{it} - \bar{x}_{i.})')^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_{i.})(y_{it} - \bar{y}_{i.}). \tag{2.1.12}$$

However, for the calculation of the estimators for α and μ_i, we need to impose the following restriction: $\sum_{i=1}^{N} \mu_i = 0$. Following that, it is easy to calculate the $\tilde{\alpha}$ and $\tilde{\mu}$:

$$\tilde{\alpha} = \bar{y}_{..} - \tilde{\beta}\bar{x}_{..} \qquad \text{and} \qquad \tilde{\mu_i} = \bar{y}_{i.} - \tilde{\alpha} - \tilde{\beta}\bar{x}_{i.}, \tag{2.1.13}$$

where $\bar{y}_{..} = \sum_{i=1}^{N}(\sum_{t=1}^{T_i} y_{it}/T_i)/N$ and similarly for the other variables.

After all, an important question lies in how we can test the existence of fixed effects in panel data. The fixed effect model is compared to the pooled model by means of an F-test with null hypothesis that all fixed individual effects μ_i are equal to 0. The test statistic is

$$F = \frac{(R^2_{pool} - R^2_{FE})/(N-1)}{R^2_{FE}/(n-N-K)} \overset{H_0}{\sim} F_{N-1,n-N-K}, \tag{2.1.14}$$

where R^2_{pool} is the residual sums of squares from the OLS estimations of the pooled model, whereas R^2_{FE} occurs from the LSDV or within regression of the FE model, K is the number of independent variables, $n = \sum_{i=1}^{N} T_i$ denoting the total number of observations in the dataset, and (N-1) and (n-N-K) are the degrees of freedom of the numerator and denominator, respectively. Note that when individual fixed effects exist, the OLS estimators of β from the pooled model are biased and inconsistent, since the fixed effects were omitted or ignored. This comes in contradiction with the unbiased and consistent FE estimators of β. Therefore, if the null hypothesis H_0 is rejected, we can conclude that the FE model is preferable to the pooled OLS.

2.1.3 Random effect model

The random effects model is defined as

$$y_{it} = \alpha + x'_{it} + \beta + \mu_i + v_{it}, \qquad i = 1, ..., N;\ t = 1, ..., Ti, \tag{2.1.15}$$

where $u_{it} = \mu_i + v_{it}$ is the error term of 2.1.15 and both the firm specific effect μ_i and the remainder disturbance v_{it} are stochastic with $\mu_i \sim IID(0, \sigma^2_\mu)$ and $v_{i,t} \sim IID(0, \sigma^2_v)$.

In addition, μ_i's, $v_{i,t}$'s and x_{it}'s are assumed to be independent of each other and among themselves [Baltagi, 2005]. This model is appropriate when the N cross-sectional units are randomly drawn from a large population and the obtained inferences regard this population. Thus, the μ_i are strictly uncorrelated with the explanatory variables [Greene, 2012].

For the computation of the variances σ^2_μ and σ^2_v, we will use the matrix notation of 2.1.15 which is

$$y = \alpha\iota_n + X\beta + u = Z\delta + u, \tag{2.1.16}$$

$$u = Z_\mu\mu + v,$$

where $n = \sum_{i=1}^{N} T_i$, y is a vector with dimension n and β of dimension K, X is an $n \times K$ matrix, Z is the matrix containing all regressors X, and the constant term α, ι_n denotes a vector of ones with dimension n and $\delta' = (\alpha', \beta')$. In addition, $Z_\mu = diag(\iota_{T_i})$, $\mu' = (\mu_1, ..., \mu_N)$ and $v' = (v_{11}, ..., v_{1T_1}, ..., v_{NT_N})$. Therefore, the variance-covariance matrix Ω for unbalanced data can be written as

$$\Omega = E(uu') = \sigma^2_\mu Z_\mu Z'_\mu + \sigma^2_v I_n, \tag{2.1.17}$$

where $Z_\mu Z'_\mu = diag(J_{T_i})$ and J_{T_i} is a T dimensional matrix of ones. From the form of the disturbance covariance matrix Ω we can obtain

$$cov(u_{it}, u_{js}) = \begin{cases} \sigma_\mu^2 + \sigma_v^2 & for\, i = j,\, t = s \\ \sigma_\mu^2 & for\, i = j,\, t \neq s \\ 0 & \text{otherwise,} \end{cases} \tag{2.1.18}$$

$$corr(u_{it}, u_{js}) = \begin{cases} 1 & for\, i = j,\, t = s \\ \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_v^2} & for\, i = j,\, t \neq s \\ 0 & \text{otherwise.} \end{cases} \tag{2.1.19}$$

[Fuller and Battese, 1974] suggested a way to obtain estimations equivalent to GLS for 2.1.15. They proposed to multiply 2.1.15 by $\sigma_v \Omega^{-1/2}$ and apply OLS on the resulting regression equation. Thus, the matrix $\Omega^{-1/2}$ is needed, and, for this reason, we will compute the spectral decomposition representation of Ω for getting the $\Omega^{-1/2}$ easier. The decomposition will be

$$\Omega = diag[(T_i\sigma_\mu^2 + \sigma_v^2)\bar{J}_{T_i} + \sigma_v^2 E_{T_i}], \tag{2.1.20}$$

where $\bar{J}_{T_i} = J_{T_i}/T_i$, $E_{T_i} = I_{T_i} - \bar{J}_{T_i}$ and I_{T_i} is the identity matrix of dimension T_i. As a result,

$$\Omega^{-1/2} = diag[(\bar{J}_{T_i}/w_i) + (E_{T_i}/\sigma_v)], \tag{2.1.21}$$

with $w_i^2 = T_i\sigma_\mu^2 + \sigma_v^2$. Following the steps of the Fuller and Battese (1974) method, 2.1.15 becomes

$$(y_{it} - \theta_i \bar{y}_i) = \alpha(1 - \theta) + (x_{it} - \theta\bar{x}_i)'\beta + \epsilon_{it}, \tag{2.1.22}$$

where $\theta_i = 1 - \frac{\sigma_v}{w_i}$, $\bar{y}_{i.} = \sum_{t=1}^{T_i}(y_{it}/T_i)$ and $\bar{x}_{i.} = \sum_{t=1}^{T_i}(x_{it}/T_i)$. The GLS estimator for δ using the true variances can be obtained from

$$\hat{\delta}_{GLS} = (Z'(\Omega)^{-1}Z)^{-1}Z'(\Omega)^{-1}y, \tag{2.1.23}$$

and the best quadratic unbiased (BQU) estimators for 2.1.22 with respect to the variances, are

$$\hat{\sigma}_v^2 = \frac{u'Qu}{tr(Q)}, \tag{2.1.24}$$

$$\hat{w}^2 = \frac{u'Pu}{tr(P)}, \tag{2.1.25}$$

with $\hat{\sigma}_\mu^2 = \frac{w_i^2 - \hat{\sigma}_v^2}{T_i}$, $Q = diag[E_{T_i}]$ and $P = diag[\bar{J}_{T_i}]$.

Unfortunately, the true disturbances are unknown, and the GLS estimators in 2.1.24 and 2.1.25 cannot be computed. A number of papers suggested different approaches in order to obtain feasible GLS (FGLS) estimations. For example, Wallace and Hussain

(1969) [Wallace and Hussain, 1969] proposed to substitute the real u with the OLS residuals $\hat{u}_{OLS}$,

$$\hat{u}_{OLS} = y - Z\hat{\delta}_{OLS} = y - Z[(Z'Z^{-1})Z'y] \tag{2.1.26}$$

whereas Amemiya (1971)[Amemiya, 1971] suggested to use the Within residuals

$$\hat{u}_w = \tilde{u} = y - \tilde{\alpha}\iota_n - X\tilde{\beta} = y - \tilde{\alpha}\iota_n - X[(X'QX)^{-1}X'Qy], \tag{2.1.27}$$

where $\tilde{\beta}$ and $\tilde{\alpha}$ are described in 2.1.12 and 2.1.13.

However, in this thesis, the [Swamy and Arora, 1972] method was used, which combines the mean square errors of two different regressions: the Within regression and Between regression. The Within regression is the same as 2.1.10, and it can be written as

$$Qy = QX\beta + Qv. \tag{2.1.28}$$

The second regression is based on the individual means, and it is given by

$$\bar{y}_i = \alpha + \bar{X}'_{i\beta} + \bar{u}_i, \tag{2.1.29}$$

which is equivalent to

$$Py = PZ\beta + Pv. \tag{2.1.30}$$

The resulted OLS residuals from 2.1.30 represents the Between residuals

$$\hat{u}^b = y - Z\hat{\delta}^b = y - (Z'PZ)^{-1}Z'Py. \tag{2.1.31}$$

Therefore, by substituting the Within residuals $\tilde{u}$ from 2.1.27 and the Between residuals $\hat{u}^b$ from 2.1.31 into the equations for the variances in 2.1.24 and 2.1.25, one gets the following estimators

$$\hat{\sigma}_v^2 = \frac{\tilde{u}'Q\tilde{u}}{n - N - K + 1}, \tag{2.1.32}$$

$$\hat{\sigma}_\mu^2 = \frac{\hat{u}^b P\hat{u}^b - (N-K)\sigma_v^2}{n - tr(Z'PZ)^{-1}Z'Z_\mu Z_{\mu'}Z)}, \tag{2.1.33}$$

where $Z'Z_\mu = diag(J_{T_i})$. Finally, the FGLS estimator of δ will result by using the $\hat{\sigma}_v^2$ and $\hat{\sigma}_\mu^2$ to construct the variance-covariance matrix Ω in 2.1.23.

[Breusch and Pagan, 1980] described a Lagrange multiplier (LM) test for balanced panel data to determine if there is a significant random effect in the data, based on the null hypothesis that the variances of the individual specific terms are zero, $H_0 : \sigma_\mu^2 = 0$. [Greene, 2012, Chapter 14] shows a modified version of this test for unbalanced data. The LM statistic is given by

$$LM = \frac{n^2}{2\sum_{i=1}^{N}(T_i(T_i - 1))}[\frac{\sum_{i=1}^{N}[(T_i\bar{e}_i)^2 - e'_i e_i]}{\sum_{i=1}^{N}(e'_i e_i)}]^2 \tag{2.1.34}$$

where ϵ_{it} is the error term of the pooled OLS regression. The test statistic follows chi-square distribution under the null hypothesis with one degree of freedom. When the null hypothesis is rejected, it is indicated that heterogeneity exists among cross-section units, and the random effects model is more appropriate than the pooled OLS.

2.1.4 Fixed or Random Effects

[Hausman, 1978] has proposed a test to examine which model among fixed effects and random effects models can treat the data better. The test is based on the correlation among the individual effects μ_i with the regressors x_{it}, because, as aforementioned, when the two variables are uncorrelated, the μ_i is randomly distributed, and thus, the random effects model is preferable. Therefore, the random effects estimator for β is consistent and efficient only under the null hypothesis of no correlation among μ_i and $x_{i,t}$ [Verbeek, 2004]. This comes in contradiction with the fixed effect estimation, which is consistent and efficient under both hypotheses. For the test statistic, the differences of the two parameter estimations are considered and also the following property is used

$$\hat{\mathbb{V}}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) = \hat{\mathbb{V}}(\hat{\beta}_{FE}) - \hat{\mathbb{V}}(\hat{\beta}_{RE}) \tag{2.1.35}$$

, with $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$ denoting the estimators for β from the fixed effects model and from the random effects model, respectively, and with $\hat{\mathbb{V}}$'s as their covariance matrices [Verbeek, 2004]. Following that, the test statistic can derive from

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\hat{\mathbb{V}}\hat{\beta}_{FE} - \hat{\mathbb{V}}\hat{\beta}_{RE}]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}). \tag{2.1.36}$$

The Hausman test statistic has a chi-square distribution with K degrees of freedom and, when the null hypothesis is rejected, one can conclude that the fixed effect model can handle better the heterogeneity across the individuals than the random effects model.

2.2 Binary response models

A discrete choice model intends to describe, explain and forecast choices among a number of discrete outcomes. In a continuous outcome, calculus is used to derive an optimal solution to the model.

Let define a set of observations of units i over time t, say $\{y_{i,t} : i = 1, ..., N; t = 1, ..., T\}$. This set is either referred to as pooled cross sectional time series data or panel data. In case the set is assumed to be either dominated by the time period or fewer units as compared to the time period length, the set is regarded as a pooled cross sectional time series data or simply *pooled data*. On the other hand, in case, the set is assumed to have observations dominated by the numbers of units over the time period. These units are (typically) a random sample – the idiosyncratic differences across individuals are not of interest: one deals with *panel data*. The key idea is that asymptotics hold as T approaches infinity as N is thought of as fixed.

The distinction between the two cases offers different ways of dealing with the statistical analysis of such sample set. Henceforth, the two approaches are discussed below.

A number of applications are translated into observations that are discrete, a contrario of continuous outcomes. Estimation of such models is usually done via parametric, semi-parametric and non-parametric maximum likelihood methods.

2.2.1 Pooled data

Let us define y_{it} as a binary choice variable taking values in $\{0, 1\}$ with a probability p_{it} of success, translating the likelihood of the event to occur. That is,

$$p_{it} := \mathbb{P}(y_{it}) = 1.$$

Following [Verbeek, 2004], the discrete variable y_t can be modelled as a linear function of indepent variables $\{x_{it} : i; t\}$ such as

$$y_{it} = x_{it}\beta + \epsilon_{it}, \tag{2.2.1}$$

where ϵ_{it} denotes the error term. With the exogeneity assumption that $\mathbb{E}(\epsilon_{it}|x_{it}) = 0$, 2.2.1 is a linear probablity model, which is, however, ill-posed as the expected value $\mathbb{E}(y_{it}|x_{it}) = x_{it}\beta \in \mathbb{R}$ may lie out of $[0, 1]$ and as error terms may be heteroskedastic since $\mathbb{V}(\epsilon_{it}|x_{it}) = x_{it}\beta(1 - x_{it}\beta)$.

Note that

$$p_{it} = \mathbb{E}(y_{i,t}|x_{i,t}) = F(x_{i,t}\beta), \tag{2.2.2}$$

for some transformation function $F(\cdot)$. Choosing $F(\cdot)$ to be a cumulative distribution function ensures that

$$F(x) \in [0, 1]\forall x,\ F(-\infty) = 0,\ F(\infty) = 1,\ \frac{\delta F}{\delta x} > 0.$$

As such, the choice of a specific cumulative distribution $F(\cdot)$ defines the models of interest that we introduce in the following.

The *probit model* uses $F(\cdot)$ as a standard normal distribution which yields

$$p_{it} = \phi(x_{it}\beta), \tag{2.2.3}$$

where $\phi(\cdot)$ denotes the standard normal distribution function. The *logit model* assumes $F(\cdot)$ to follow a standard logistic distribution so that the choice probability takes the form of

$$p_{it} = \frac{exp(x_{it}\beta)}{1 + exp(x_{it}\beta)} = \frac{1}{1 + exp(-x_{it}\beta)}. \tag{2.2.4}$$

Another representation of the model is the underlying latent model. Consider the latent variable $y^{\star}_{it}$ as not observed and defined per $y_{it} := \mathbf{1}_{y^{\star}_{it}>0}$,

$$y^{\star}_{it} = x_{i,t}\beta + u_{it}, \mathbb{E}(u_{it}) = 0, \tag{2.2.5}$$

where the error terms u_{it} are assumed to be uncorrelated across individuals, see [Baltagi, 2005, Greene, 2012]. In addition, it is assumed that individual observations (y_i, x_i) are iid, explanatory variables are exogenous and error terms are normally distributed and homoskedastic.

The models described above are adequate for pooled data. However, a core assumption of pooled data is that individuals are homogeneous and time-series and cross-section analyses with missing controlling individual heterogeneity run the risk of obtaining biased estimates. To this matter, panel data modelling is a way of controlling individual heterogeneity, by introducing an idiosyncratic compnent μ_i. The variable μ_i captures the unobservable individual effects and, based on its relation among the explanatory variables of the regression model, one can discriminates a so-called *fixed effect* and *random effect* model.

The next sections explain how to model panel data with fixed effects, and how to handle random effects. Then, a criterion to approach the best model is described and, in the last section, appropriate tests for misspecification of the model variables are introduced.

2.2.2 Fixed effect model

The fixed effect model can be defined as following:

$$y^{\star}_{it} = \mu_i + x_{it}\beta + v_{it},\ i = 1, ..., N,\ t = 1, ...T_i, \tag{2.2.6}$$

with

$$\mathbb{P}(y_{it} = 1) = Pr(y^{\star}_{it} > 0) = Pr(v_{it} > -\mu_i - x_{it}\beta) = F(\mu_i + x_{it}\beta)$$

where $F(.)$ denotes a distribution function which is symmetric around zero and the explanatory variables $x_{i,t}$ are independent from each other. The variables v_{it} express the disturbance term for every individual i at time t, following the distribution with a cumulative distribution function $F(.)$ and which is homoskedastic. This model is appropriate when there is an individual-specific unobserved effect μ_i in the data, which is considered as a fixed unknown parameter and there is no restriction on its relation among the explanatory variables x_{it} [Baltagi, 2005].

The usual method of estimating the parameters of interest in a binary choice model, in this case the unobserved effects μ_i and regressor coefficient β, is the computation of their maximum likelihood estimators (thereafter, MLE). The log-likelihood function of this model is

$$\log Ł(\beta, \mu_1, ..., \mu_N) := \sum_{it} \log Pr(y_{it}|\mu_i + x_{it}\beta),$$

, which is equal to

$$\sum_{it} \log F(\mu_i + x_{it}\beta) + \sum_{it}(1 - y_{it}) \log\big(1 - F(\mu_i + x_{it}\beta)\big). \tag{2.2.7}$$

Differentiating this function will result to the score functions $s(\beta)$ and $s(\mu)$ of the parameters β and μ, and the MLE can be obtained from these functions. These estimators are consistent when T_i goes to infinity [Verbeek, 2004].

However, this method has a number of deficiencies [Greene, 2007]. First, the parameter estimators of the time invariant variables (e.g. sex, race or religion) cannot be obtained, as in the linear model. Another shortcoming for this method arises when the number of individuals N in the dataset is large, since it would be probably difficult to derive all individual fixed effects estimators μ_i by maximizing 2.2.7. This problem has a straightforward solution, according to [Greene, 2004a, Greene, 2004b], for solving iteratively the resulting system. At the same time, as stated by [Baltagi, 2005, Greene, 2012, Verbeek, 2004], the increase of the number of μ_i while $N \rightarrow \infty$ arise inconsistent parameter estimators when T_i is fixed. This is the incidental parameters problem [Neyman and Scott, 1948, Lancaster, 2000]. [Abrevaya, 1997] illustrates the existence of upward bias in the MLE estimator of β in the context of a panel logit model with $T = 2$. In more details, he proved that as $N \rightarrow \infty$, plim$\hat{\beta} = 2\beta$. Monte Carlo simulations are performed by [Greene, 2004b] on a panel probit model with N=1000, showing that the bias persists for even larger T, e.g. $T = 10$ and $T = 20$. Also, the observations for individuals with $\sum\limits_{i=1}^{T_i} y_{it} = 0$ or $\sum\limits_{i=1}^{T_i} y_{it} = 1$ are not included in the estimation because they do not affect the log-likelihood function 2.2.7. This results from the fact that for such individuals the $\hat{\mu_i}_{MLE}$ is infinite, see [Chamberlain, 1980]. Such problem is known as the "perfect prediction problem" [Maddala, 1986].

In the case of the linear model, which was discussed in Section 1.1.2, we have introduced the "within transformation" of the model, where for T_i fixed we could discard the idiosyncratic constant μ_i before estimate β and get consistent estimates for β [Hsiao, 2003]. This becomes possible when using deviations from group means. However, for most probability models, the inconsistency of μ_i is transferred to β as well, since their estimators are dependent [Baltagi, 2005]. Even by converting the y_{it} to deviations, like $y_{it} - y_{i,t-1}$, removing μ_i, will yield a variable with unknown characteristics [Verbeek, 2004, Greene, 2012].

For the binomial panel fixed effect model, [Chamberlain, 1980] suggests to discard the individual fixed effects by conditioning on the minimum sufficient statistic for μ_i, that is, $\sum\limits_{t=1}^{T_i} y_{i,t}$. Consequently, the derived function that needs to be maximised is the conditional likelihood function

$$\log\ L(\beta, \mu_1, ..., \mu_N) = \sum_{i,t} \log Pr(y_{i,t}|\mu_i + x_{i,t}\beta, \sum_{t=1}^{T_i} y_{i,t}. \tag{2.2.8}$$

This strategy is possible for the logit model, but there are not sufficient statistics for the probit model [Greene, 2012]. For a logit model we have

$$Pr(y_{it} = 1) = \frac{e^{\mu_i + x_{it}\beta}}{1 + e^{\mu_i + x_{it}\beta}}, i = 1, ..., N,\ t = 1, ..., T_i$$

, and, therefore, from 2.2.8, the conditional maximum likelihood for $T_i = 2$ yields

$$\log L = \log\left(\frac{1}{1+e^{(x_{i2}+x_{i1})\beta}}\right) + \log\left(\frac{e^{(x_{i2}+x_{i1})\beta}}{1+e^{(x_{i2}+x_{i1})\beta}}\right).$$

The individuals who do not switch status from 0 to 1 and 1 to 0 are excluded in estimation, as happened in the unconditional case [Baltagi, 2005].

The conditional maximum likelihood function (9) is free of the incidental parameters, μ_i . Hence, we get a consistent conditional logit estimator for β, $\hat{\beta_{CMLE}}$ without estimating μ_i. For applications where the Chamberlain conditional estimator was used, see [Björklund, 1985, Cecchetti, 1986, Willis, 2006]. Note that, by discarding μ_i, we made it impossible to derive an estimation for them based on the conditional log-likelihood. A possible solution is proposed by [Greene, 2004a]: that is to compute on a second step estimates for μ_i by maximizing the unconditional log likelihood function 2.2.7 with respect to μ_i, using the consistent estimator of β, $\hat{\beta_{CMLE}}$. However, the resulting estimator for μ_i will be inconsistent, as its variance will be biased and no solution exists when $\sum y_{it}$ is either 0 or 1. Overall, using the Chamberlain (1980) [Chamberlain, 1980] approach, the incidental parameters problem is addressed; nonetheless, the perfect prediction problem remains.

Other papers have considered different approaches of obtaining consistent estimators for β without the fixed-T assumption. An approach includes removing the first order bias in β_{MLE} , as in [Hahn and Kuersteiner, 2002, Hahn and Newey, 2004, Fernández-Val, 2009, Dhaene and Jochmans, 2015]. On the other hand, [Bester and Hansen, 2009] suggests an ex-ante correction method using a modified objective function. However, these methods only solve the incidental parameters problem and not the perfect prediction. [Kunz et al., 2017] introduced a bias reduced (BR) estimator for a binary response panel model with a fixed T. Their estimator was based on the idea of [Firth, 1993, Kosmidis and Firth, 2009] to get a biased estimator for β in linear exponential family models for cross-section data. They proposed to deduct the first-order bias of the β_{MLE} from the score function $s(\beta)$, resulting in a modified score function $\tilde{s}(\beta)$. [Kunz et al., 2017] extended this approach to the panel data case, where they showed analytically that the BR estimator always produces finite estimates with regard to the fixed effects. Moreover, using Monte Carlo simulations in a probit model for an unbalanced panel over a five-year period, they illustrated that their BR estimator gives reliable and bias-reducing estimates of β as well as better performance compared to other proposals estimators, including the MLE. In their paper, they suggested the use of an adjusted response variable $\tilde{y_{it}}$ (pseudo-responses) defined as

$$\tilde{y_{it}} := y_{it} + \frac{1}{2}h_{it}\frac{f'_{it}}{w_{it}}, \tag{2.2.9}$$

where $f_{it} = f(\mu_i + x_{it}\beta)$ abd h_{it} denotes the i-th diagonal entry of the projection matrix H of dimension $NT \times NT$

$$H := W^{1/2}X(X'WX)^{-1}W^{1/2}$$

, with X being the matrix of the NT observations on K independent variables and W the $NT \times NT$ diagonal matrix with entries $w_{it} = \dfrac{f_{it}^2}{F_{it}(1-F_{it})}$. Therefore, by substituting the pseudo-responses for responses y_{it} into the log-likelihood function 2.2.7 and maximizing it with respect to β_k and μ_i, separately, the modified score functions for β_k and μ_i are obtained. That is,

$$\tilde{s}(\beta_k) = s^{BR} = s(\beta_k) + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i} h_{it}\frac{f'_{it}}{w_{it}}x_{kit}, \tag{2.2.10}$$

$$\tilde{s}(\mu_i) = s^{BR}(\mu_i) = s(\mu_i) + \frac{1}{2}\sum_{t=1}^{T_i} h_{it}\frac{f'_{it}}{w_{it}} \tag{2.2.11}$$

where $s(\beta_k)$ and $s(\mu_i)$ are the standard MLE score functions from 2.2.7. The $\hat{\beta}_k$ and $\hat{\mu}_i$ are calculated simultaneously via using an iteratively re-weighted least squares (IWLS) procedure [Kosmidis and Firth, 2009]. In this thesis, the BR estimator was developed in the context of a logit panel model. According to [Kunz et al., 2017] , the equations 2.2.9-2.2.11 will be

$$\tilde{y}_{i,t} = y_{i,t} + h_{i,t}(\frac{1}{2}\Lambda_{i,t})$$

$$\tilde{s}(\beta_k) = s^{BR}(\beta_k) = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\Big(y_{i,t} - \Lambda_{i,t} + h_{i,t}(\frac{1}{2} - \Lambda_{i,t})\Big)x_{k,i,t},$$

$$\tilde{s}(\mu_i) = s^{BR}(\mu_i) = \sum_{t=1}^{T_i}\Big(y_{it} - \Lambda_{it} + h_{it}(\frac{1}{2} - \Lambda_{it})\Big),$$

where h_{it} is based on the values of $w_{it} = \Lambda_{it}(1-\Lambda_{it})$. Following that, [Baltagi, 2005, Greene, 2012, Verbeek, 2004] mentioned in their works the importance of whether there is homogeneity, i.e. $\mu_i \equiv \mu$, in the model and, therefore, no individual fixed effects or if there is heterogeneity. They proposed a Hausman-type test as a means to compare the pooled logit MLE with the Chamberlain's conditional MLE (CMLE), for which [Hsiao, 2003] gives proof. The null hypothesis of the test is the absence of individual effects and, when this holds, both estimators are consistent and efficient. Under the alternative hypothesis, the pooled MLE is inconsistent, since the fixed effects are ignored, whereas it is possible that the CMLE does not use all data, which makes it inconsistent. For the case of BR estimator, the Hausman test [Hausman, 1978] can be also used since this estimator is consistent and efficient under both null and alternative hypothesis. The Hausman test statistic is

$$\xi_H := (\hat{\beta}_{BR} - \hat{\beta}_{MLE})[\hat{V}(\hat{\beta}_{BR}) - \hat{V}(\hat{\beta}_{MLE})]^{-1}(\hat{\beta}_{BR} - \hat{\beta}_{MLE}), \tag{2.2.12}$$

where $\hat{V}$ are the covariance matrices. This statistic follows a chi-squared distribution under the null hypothesis with K degrees of freedom, where K is the dimension of β without the intercept term of the model. If $\hat{V}(\hat{\beta}_{BR})$ is larger than $\hat{V}(\hat{\beta}_{BR})$, their difference is assumed to be zero and, consequently, the Hausman test statistic is zero.

2.2.3 Random effect model

The random effects model is represented by

$$y_{it}^{\star} = \mu_i + x_{it}\beta + v_{it},\ i = 1, ..., N,\ t = 1, ...T_i,$$

where $\mathbb{P}(y_{it} = 1) = F(x_{it}\beta)$. The individual specific effect μ_i in this model is a random variable while $\mu_i \sim iid(0, \sigma_\mu^2)$ and $v_{it} \sim iid(0, \sigma_v^2)$ are independent of each other and of x_{it} for all i and t. The explanatory variables are exogenous, while the error terms follow the distribution of the cumulative function $F(.)$and are homoskedastic. Also for $t \neq s$, one has $\mathbb{E}(\mu_{it}v_{it}) = \sigma_\mu^2$. As a result, the error terms $v_{i,t}$ from different time periods t depend on their correlation with $\dfrac{\sigma_\mu^2}{\sigma_v^2 + \sigma_\mu^2}$ (see [Baltagi, 2005] and [Greene, 2012]). As mentioned earlier, a common method of computing β in a binary choice model is by maximising the log-likelihood function in order to obtain the MLE. The likelihood function for the individual i is the joint probability across all T_i observations and for a random effects model is

$$L_i = Pr(y_{i,1}, ...,_{i,T_i} |X) = \int ... \int f(u_{i,1}, ...,_{i,T_i})du_{i,1}...d_{i,T_i}. \tag{2.2.13}$$

These T_i integrals are not independent, thus the computation of the MLE based on T_i-dimensional integrals is hard and not feasible as soon as $T_i \geqslant 4$. To overcome this problem, one can condition the joint density of $(u_{i,1}, ..., u_{i,T_i})$ upon μ_i and get independent error terms $u_{i,t}$. Hence, the joint density of $(u_{i,1}, ..., u_{i,T_i})$ is

$$f(u_{i,1}, ..., u_{i,T_i}) = \int_{-\infty}^{+\infty} \prod_{t=1}^{T_i} \int_{L_{i,t}}^{U_{i,t}} [f(u_{i,t}|\mu_i)du_{i,t}]f(\mu_i)d\mu_i, \tag{2.2.14}$$

where the individual probability density functions are

$$\int_{L_{i,t}}^{U_{i,t}} f(u_{i,t}|\mu_i)du_{i,t} = Pr(y_{i,t}|\mu_i + x_{i,t}\beta) \tag{2.2.15}$$

After these steps, the L_i can be derived from T_i one-dimensional integrations. An assumption on the distributions of $v_{i,t}$ and μ_i is left to be made. It is better to consider the same distribution for them, in order to avoid nonstandard distributions for $\mu_i + v$. As a first approach, someone can consider the logistic or normal distribution, as they are the most common in practice. Unfortunately, a multivariate logistic distribution for $v_{i,1}, ..., v_{i,T_i}$ will not be the best choice, due to its property of having all correlations equal to $\frac{1}{2}$ (for more details see [Maddala, 1986]). However, the multivariate normal distribution appears to be perfect for this method. In more details, we can assume that

$$\mathbb{E}(\mu_{i,t}|X) = 0,\ \mathbb{V}(\mu_{i,t}|X) = 1,\ Cov(u_{i,t}, u_{i,s}|X) = \sigma_\mu^2\ fors \neq t.$$

and

$$\mu_i \sim iid(0, \sigma_\mu^2),\ v_{i,t} \sim iid(0, 1-\sigma_\mu^2)$$

Having these distributions, we deduce to the random effects probit model for panel data, where the distributions functions that are needed for the likelihood function 2.2.14 and 2.2.15 are:

$$\mathbb{P}(y_{i,t}|\mu_i + x_{i,t}\beta) = \begin{cases} \phi\left(\dfrac{\mu_i + x_{i,t}\beta}{\sqrt{1-\sigma_\mu^2}}\right), & if\ y_{i,t} = 1 \\ 1 - \phi\left(\dfrac{\mu_i + x_{i,t}\beta}{\sqrt{1-\sigma_\mu^2}}\right), & if\ y_{i,t} = 0 \end{cases}, \tag{2.2.16}$$

$$f(\mu_i) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\big(-\frac{\mu_i^2}{2\sigma_\mu^2}\big), \tag{2.2.17}$$

where $\phi(.)$ denotes cumulative distribution function of a standard normal variable. Butler and Moffit (1982) [Butler and Moffitt, 1982] outline an algorithm to compute (23), assuming normality distribution for the individual effects μ_i, where Gaussian quadrature procedures are used. For a detailed description of this approach consult [Butler and Moffitt, 1982]. Moreover, according to [Greene, 2012], if individual random effects exist in the data and are ignored, the MLE for β obtained from a pooled probit will be inconsistent. For that reason, it is crucial to examine whether this kind of effects exists in the data and decide what model is more appropriate for them. [Greene, 2012] advocates the use of the likelihood ratio (LR) among the pooled and random effects pooled model with $H0 : \sigma_\mu^2 = 0$, and test statistic

$$LR = -2(\log\hat{L}_R - \log\hat{L}_U), \tag{2.2.18}$$

where $\log\hat{L}_R$ and $\log\hat{L}_U$ indicate the log-likelihood values of the restricted model, in that case, the pooled model and the unrestricted random effects model, respectively. The statistic LR follows Chi-square distribution with one degree of freedom. Instead of this test, the Wald test and Lagrange multiplier (LM) test can be used (see [Greene, 2012]).

2.2.4 Fixed or Random Effects

In panel data arises the question of whether fixed individual effects are more appropriate for the data rather than random effects. In the concept of binary data, we can distinguish between these two models by using a third model [Mundlak, 1978, Chamberlain, 1984, Wooldridge, 2002]. The model assumes that the individual effects μ_i are linearly dependent to the individual means $\bar{x}_i$ for all time varying regressors $x_{i.t}$.

$$\mu_i = \mu + \bar{x}_i\delta + e_i, \tag{2.2.19}$$

where $e_i \sim IIN(0, \sigma_e^2)$ represent the errot term and δ is the coefficient of the individuals means $\bar{x}_i$. Wooldridge and Chamberlain propose in their analyses to calculate the means of xit over all observations (all individuals and years) instead of individually means. However, this will create a complication in the unbalanced panels. By adding 2.2.19 to the fixed effects regression model 2.2.6, we got this random effects formulation

$$y^{\star}_{i,t} = \mu_i + \bar{x}_i\delta + x'_{i,t}\beta + e_i + v_{i,t},\ i = 1, ..., N,\ t = 1, ..., T_i, \tag{2.2.20}$$

with $Pr(y_{i,t} = 1) = F(\mu + \bar{x}'_i\delta + x'_{i,t}\beta + e_i)$.

Mundlak [Mundlak, 1978] shows that if $\delta \neq 0$, the individual effects μ_i are correlated with the regressors $x_{i,t}$ of the model, and the model 2.2.20 is a random effects model which deals with the problem of the aforementioned relation. At the same time, a pure random effects model will be derived if $\delta = 0$. Therefore, by comparing these two models, we can draw conclusions on the dependence between μ and $x_{i,t}$ and, consequently, on the existence of fixed or random effects. A Wald test can be applied here with null hypothesis of $\delta = 0$, which is the hypothesis of the random effects model [Greene, 2012]. Ifμ_i and $x_{i,t}$ are correlated but a pure random effects model is used, which ignores that association, then, the resulting estimator will be biased. The Wald test statistic is

$$W = \hat{\delta}'\hat{V}(\hat{\delta})^{-1}\hat{\delta}, \tag{2.2.21}$$

where $\hat{V}(\hat{\delta})$ denotes the covariance matrix of $\hat{\delta}$ and follows a chi-square distribution with N degrees of freedom.

3

Data analysis in credit risk modelling

Contents

This chapter discusses two approaches in credit risk modelling on a mortgage portfolio. For sake of compliance, a number of features related to the models can not been disclosed e.g. client-specific data, levels of model outcomes and driver description have been removed or anonymized.

In this chapter, we study different approaches to model the LGD for mortgage assets. As mentioned, the Loss Given Default represents the percentage of the Exposure at Default (EAD) which is expected to incur a loss from a default event. There are a number of ways to calculate loss given default via e.g. probability of recovery, loss averaging derivation, etc. We chose to model the LGD by defining two events: cure and no-cure, so that

$$LGD = (1 - Cure\ rate) \cdot LGN,$$

where the *Cure rate* is the percentage of the defaulted counterparties curing from default and the Loss given no-cure (thereafter, *LGN*) is the loss for non-cured defaulted loans. Therefore, the LGD model relies on two underlying models.

The chapter is divided into three sections. Section 1 and Section 2 deal with the two underlying LGD submodels, respectively, on the so-called *Cure rate* and *LGN*.

Therein, pooled and panel data regression analysis are performed for the two models as well as performance backtesting tests. Then, Section 3 implements the resulting pooled and panel LGD models and compares their performances. The structure of this chapter is illustrated in Figure 3.1.

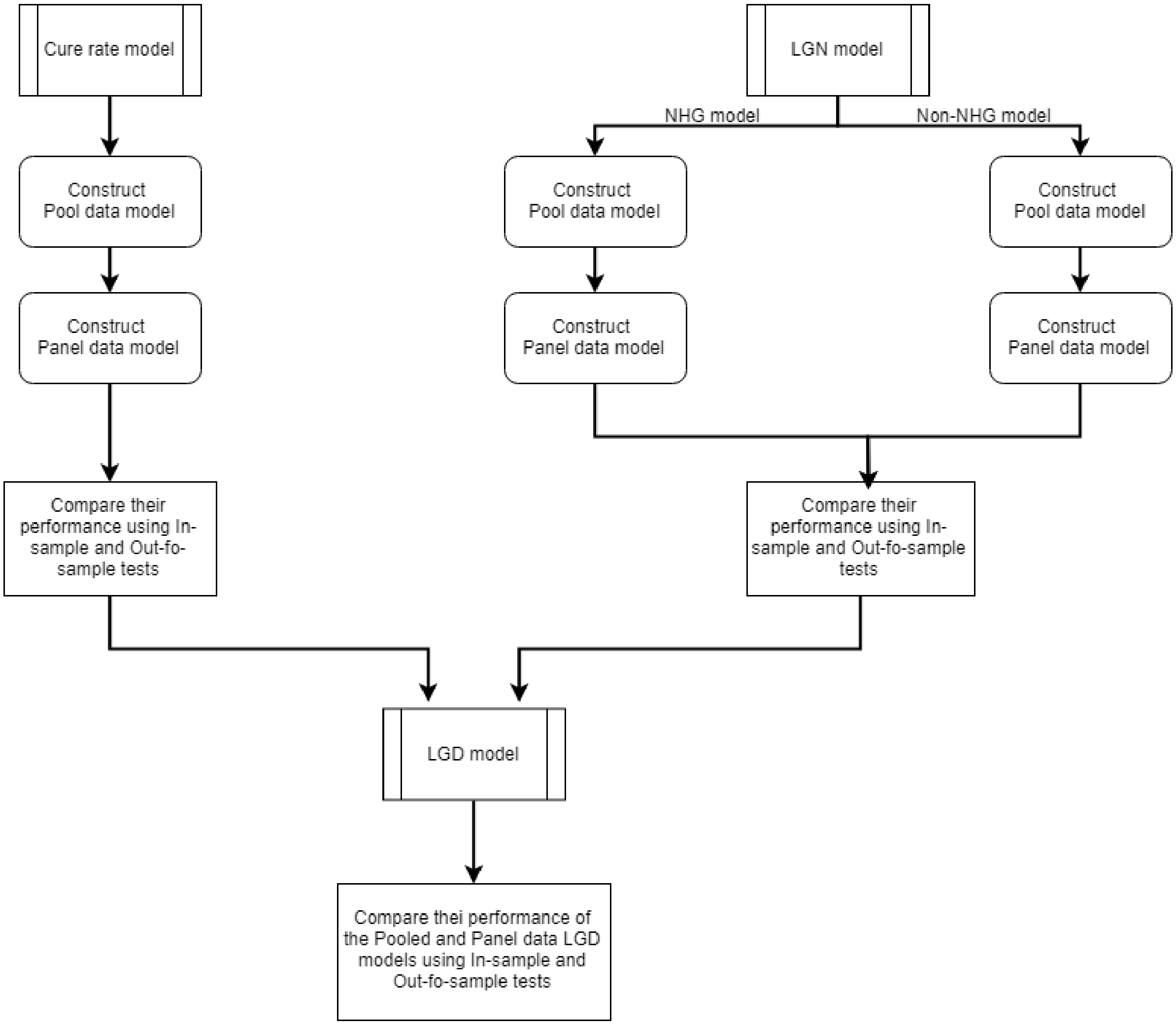

Cure rate model
LGN model
NHG model
Non-NHG model
Construct Pool data model
Construct Pool data model
Construct Pool data model
Construct Panel data model
Construct Panel data model
Construct Panel data model
Compare their performance using In-sample and Out-fo-sample tests
Compare their performance using In-sample and Out-fo-sample tests
LGD model
Compare thei performance of the Pooled and Panel data LGD models using In-sample and Out-fo-sample tests

Figure 3.1: Structure of the chapter.

A performing counterparty is said to be defaulted according to [BCBS, 2006] or is considered unlikely to pay its debt or fails to pay in time its financial obligation due to the bank within the contractual defined period. Notwithstanding the duration of this defined period, a counterparty which is past due more than 90 days on any financial obligation to the bank, is considered to be in default.

The portfolio with all the defaulted clients gives information about the status of every contract. If it is indicated as “cure” then the counterparty returns to the performing portfolio, otherwise is qualified as “uncured”. In the last case, either the debt is considered a loss (write off) or the contract is recovered without loss for the bank (paid-off).

The date that the “cure” or “uncure” is assigned is the client’s default end-date (default is closed) and the year of the default end-date is called outflow year.

The LGD model was developed for rating both healthy and defaulted clients which belong to the performing and already defaulted portfolio, respectively. Consequently, the selection of the scope years for each portfolio is important in order to be aligned the contracts of the two datasets. The selection of the defaulted dataset must align with the performing one. For retail exposures, the estimates should be based on a business cycle of data [BCBS, 2006].

In this thesis, the LGD model that will be developed, it is based on a 5-year dataset which contains all the contracts that were in default during these years. Each of these contracts is observed at most 5 times based on whether the mortgage loan has cured and left the defaulted portfolio, or not. Therefore, this is an unbalanced panel dataset, where contracts represent the cross-section dimension and years denote the time-series dimension.

This dataset contains all the credit characteristics of mortgage loans at client level. For example, the Loan to market value ($LTMV$) variable is contained in this dataset denoting the ratio of the contract’s total debt to the market value of the collateral. Another variable, is the one that counts the number of payment terms in arrears for every defaulted contract or the one that indicates if the defaulted contract is cured and left this dataset or not. For confidentiality issues we do not enter into more details about the dataset’s variables. For that reason, in the rest of this thesis we will not refer to these variables with their real names, but instead we will call them $Driver1$, $Driver2$, etc. The full analysis and description of these drivers is disclosed only to the thesis committee.

3.1 The Cure rate model

The *Cure rate* defaulted model was constructed first, in order to understand the probability of a loan to stand out of a defaulted status (in a regulatory sense).

3.1.1 Model construction

For the development of this model the entire 5-years dataset was used. Bivariate analysis was performed to assess the relationship among the response variable and the possible regressors. The dependence regarding the different continuous and categorical regressors was measured by means of a Chi-square test, and, the strength of dependence among all the variables was assessed, using a Kendall tau correlation coefficients. Considering the results of these tests, the explanatory variable x_1 for the *Cure rate* model was chosen. The results of these tests are confidential and therefore were removed from this report. Hence, here we will call the explanatory variable $Driver1$.

The *Cure rate* being a probability, one defines the following pooled and panel data models, respectively: For a unit $i \in \{1, ..., N\}$ and time period $t \in \{1, ..., T\}$,

$$Cure\ rate_{it} = \Pr\big(\alpha + \beta_1 \cdot Driver1_{it} + \epsilon_{it} > 0\big), \tag{3.1.1}$$

where $Driver1$ is the independent variable, ϵ_{it} is the regressor error term and coefficients $\alpha, \beta_1 \in \mathbb{R}$, and

$$Cure\ rate_{it} = \Pr\big(\alpha + \mu_i + \beta_1 \cdot Driver1_{it} + v_{it} > 0\big), \tag{3.1.2}$$

where μ_i being the *unobservable* effect for each unit i and v_{it} being the remainder disturbance term of the model. The coeficients α and β_1 are again real numbers.

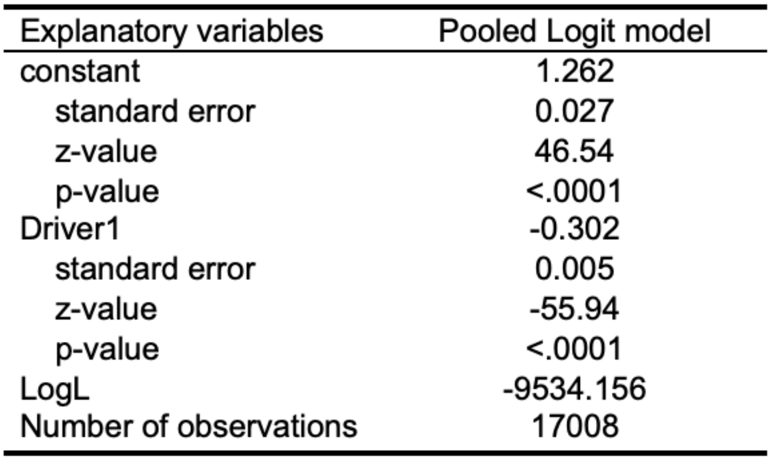
2.1.1.a Pool data modelling

The performance of the panel data model will be benchmarked against a pooled data model, as introduced per Chapter 2. In order to model a probability outcome, one benchmark the cure rate to a logit model which is free of the probit model assumption of normal errors. Under the same notations,

$$Cure\ rate_{it} = \Lambda\big(\alpha + \beta_1 \cdot Driver1_{it}\big) = \frac{1}{1 + \exp -(\alpha + \beta_1 \cdot Driver1_{it})}. \tag{3.1.3}$$

The resulted maximum likelihood estimators for this logistic model on the pooled data are presented in Table 3.1. One can see that the intercept and the coefficient of $Driver1$ are both significant at 1% level. Moreover, the regression coefficient of the input variable is -0.30, underlying the negative relation of the $Driver1$ with the probability for the client to cure, as expected.

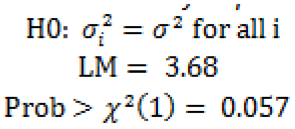
Table 3.1: Regression analysis results for Pooled logistic model



Explanatory variables	Pooled Logit model
constant	1.262
standard error	0.027
z-value	46.54
p-value	<.0001
Driver1	-0.302
standard error	0.005
z-value	-55.94
p-value	<.0001
LogL	-9534.156
Number of observations	17008

According to [Verbeek, 2004], for a logit model it is important to be examined for heteroscedastic errors and omitted variables, as these will result to incorrectly specified likelihood function and inconsistent estimators. An appropriate framework for testing these assumptions is the Lagrange multiplier (LM) test, with null hypothesis of homoscedastic regression errors and no omitted variables, respectively. However, the *Cure rate pooled* model was not studied regarding any omitted variables, since the model's variables were already chosen from the Bank. To examine whether the homoscedasticity assumption holds, the quadratic term of the chosen Driver1 was included in the regression model. Details about this LM test can be found in Appendix A. The results of the LM test are presented in Table 3.2. The test statistic is 3.68 and is lower than the chi-square critical value with one degree of freedom, which is 3.84. As a consequence, we will not reject the null hypothesis of having constant residual's variance in the model.

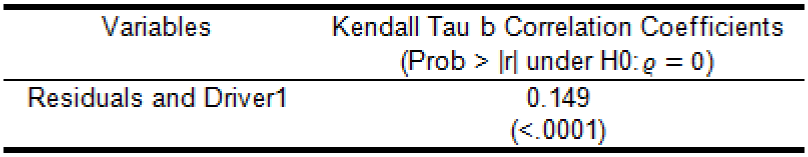
Table 3.2: LM test for heteroscedasticity in pooled logistic model



$$\text{H0: } \sigma_i^2 = \sigma^2 \text{ for all i}$$
$$\text{LM} = 3.68$$
$$\text{Prob} > \chi^2(1) = 0.057$$

Then, correlation analysis was conducted among Pearson regression residuals and the predictor variable, in order to examine the weak exogeneity assumption for the logit model. The Kendall coefficients [Chok, 2010] in Table 3.3 point out the low dependency between them, since the correlation coefficient with value 0.149 is smaller than 0.50.

Table 3.3: Kendall rank correlation analysis among regression Pearson residuals and the explanatory variable for pooled logistic model



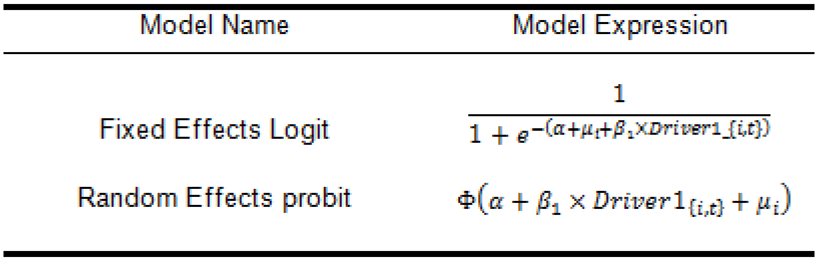
Variables	Kendall Tau b Correlation Coefficients (Prob > |r| under H0: $\varrho = 0$)
Residuals and Driver1	0.149 (<.0001)

2.1.1.b Panel data modelling

As mentioned in Chapter 2, when modelling with panel data it is necessary to develop two different models, one with fixed unobserved individual effects and one with random effects. Upon which, one can decide which model is the most appropriate for our data.

First, we focus on the Fixed effects logit model using the Firth's biased reduction method as [Kunz et al., 2017] proposed, where we assume that the error terms v_{it} follow a standard logistic distribution. The maximum likelihood estimators (MLE) were obtained, after the first-order bias was removed from the score functions of the unknown parameters. The iteratively re-weighted least squares (IWLS) algorithm [Kosmidis and Firth, 2009] was used, as described in Chapter 2. In this model no assumption was made regarding the relation among client specific effects μ_i and regressor x_1. However, for the next model we assumed that the two previously mentioned regressor terms are uncorrelated and also μ_i and v_{it} are both normally distributed [Verbeek, 2004]. The last described model is the Random effect logit model, where the MLE's were derived using the conditional joint density of $(u_{i1}, ..., u_{iT_i}, \mu_i)$ upon μ_i.

Table 3.4: Two competing panel data models

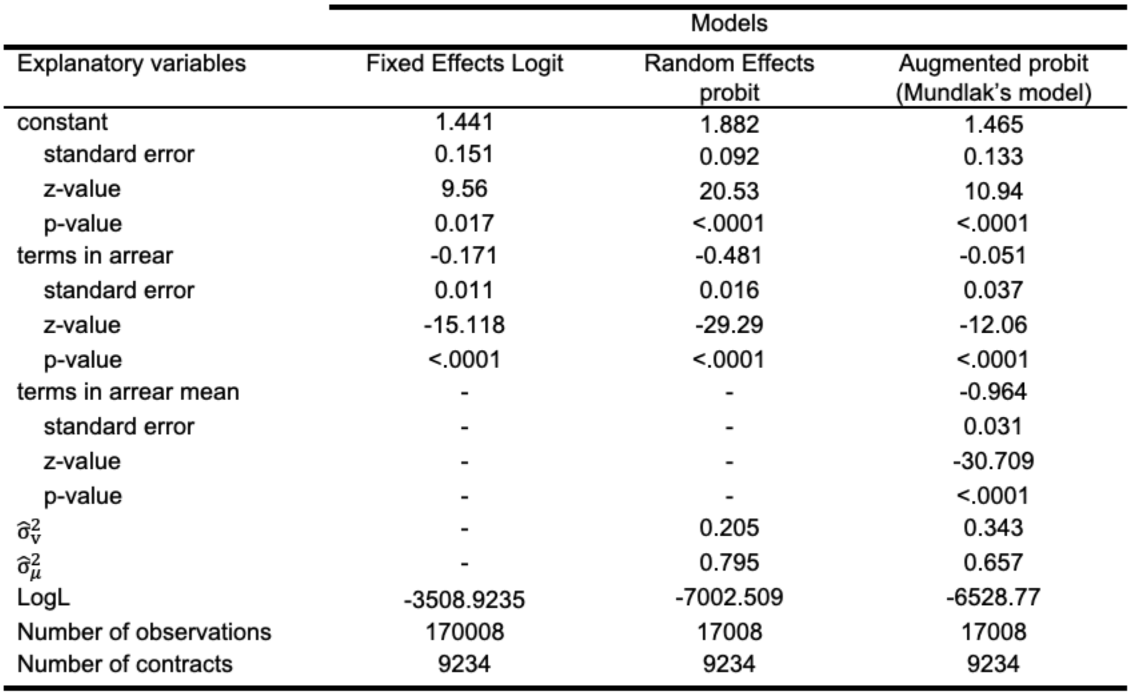


Model Name	Model Expression
Fixed Effects Logit	$\frac{1}{1+e^{-(\alpha+\mu_i+\beta_1 \times Driver1\_\{i,t\})}}$
Random Effects probit	$\Phi\left(\alpha + \beta_1 \times Driver1_{\{i,t\}} + \mu_i\right)$

Table 3.4 presents the form of these two binary response models and their output estimators are given in Table 3.5. For both models the regression coefficient of Driver1 is significant and negatively signed as in the pooled model. If fixed effects per unit is considered, the estimator reduces to -0.171, while for the pooled and random model is -0.302 and -0.481, respectively. The three models capture almost the same relation among the response variable and the regressor, as evidenced by the little difference among the magnitude of the three regressor coefficients. Additionally, one can observe that the Fixed effect logit model scores the highest log-likelihood (LogL) value, but according to [Greene, 2012] this is not a fit measure so that the model choice cannot be chosen upon.

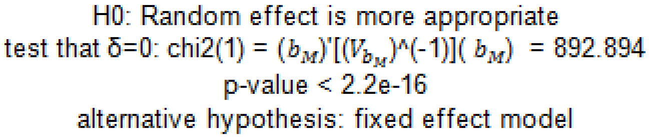
The decision for the most appropriate model to represent our data was taken after two different statistical tests were employed. First, we considered the Mundlak's approach [Mundlak, 1978] which is described in Chapter 2, and suggests to assume that μ_i has a linear relation with the individual means of all the time varying regressors $Driver1_{it}$. In our case, the predictor variable Driver1 is not constant across time so its averages were computed. After adding these individual means to the regression model, an augmented binary choice random effects model was derived. For that model the standard normal distribution function was chosen, since it will be compared among another probit model. Table 3.5 shows the "pure" random effects probit model in the second column and the augmented one in the third column. Comparing the estimators of the variable Driver1 from the two models we suspect that the random effects model is not the preffered one, since these estimators are very different. By including the individual means in the model the resulting estimators decreases to -0.051, whereas for the "pure" Random effects model is -0.481. A Wald test was carried out, testing the null hypothesis that the coefficient of the extra regressor in the augmented model is null [Greene, 2012]. The resulting Wald statistic in Table 3.6 indicates that the null hypothesis of the random effects, and, exogeneity among the regressors and the individual effects, is rejected [Baltagi, 2005]. Thus, a Fixed effect model is deemed more appropriate to model the data over a Random effect model.

Table 3.5: Regression analysis results for Fixed effects logit and Random effects probit models



	Models		
Explanatory variables	Fixed Effects Logit	Random Effects probit	Augmented probit (Mundlak's model)
constant	1.441	1.882	1.465
standard error	0.151	0.092	0.133
z-value	9.56	20.53	10.94
p-value	0.017	<.0001	<.0001
terms in arrear	-0.171	-0.481	-0.051
standard error	0.011	0.016	0.037
z-value	-15.118	-29.29	-12.06
p-value	<.0001	<.0001	<.0001
terms in arrear mean	-	-	-0.964
standard error	-	-	0.031
z-value	-	-	-30.709
p-value	-	-	<.0001
$\hat{\sigma}_v^2$	-	0.205	0.343
$\hat{\sigma}_\mu^2$	-	0.795	0.657
LogL	-3508.9235	-7002.509	-6528.77
Number of observations	170008	17008	17008
Number of contracts	9234	9234	9234

Table 3.6: Wald test for choosing Fixed effects logit or Random effects probit model

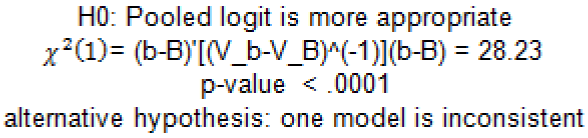


H0: Random effect is more appropriate
test that δ=0: chi2(1) = (b_M)'[(V_{b_M})^(-1)](b_M) = 892.894
p-value < 2.2e-16
alternative hypothesis: fixed effect model

Following, a Hausman test [Hausman, 1978] was performed based on the differences between the fixed effect biased reduction logit MLE and the usual logit MLE, in order to determine which model is preferred. The latter estimator ignores fixed effects and will be inconsistent under the null hypothesis of client specific effects existence, as [Baltagi, 2005] explains. The Chi-sqaure value of the test is given in Table 3.7 and is 28.23, large enough to reject the null hypothesis. Consequently, the Fixed effects model appears to be more suitable for the data.

The fixed effect logit model is based on the same assumptions of constant errors variance, no omitted variables and independence among regression residuals and predictors as the pooled logit model [Verbeek, 2004]. However, the Bank's modelling team has already selected the final variable for the model, and as a consequence the LM test for omitted variables was not employed. For the variance of residuals, a LM test statistic was calculated, which is explained in Appendix A. The resulting value of this statistic

Table 3.7: Hausman test for choosing Fixed effects logit or Pooled logit model



H0: Pooled logit is more appropriate

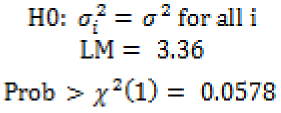
$\chi^2(1)$= (b-B)'[(V_b-V_B)^(-1)](b-B) = 28.23

p-value < .0001

alternative hypothesis: one model is inconsistent

is displayed in Table 3.8 and its low magnitude indicates that we cannot reject the null hypothesis and the appearance of homoskedastic errors in the regression model. As regards the third assumption, a Kendall correlation analysis [Chok, 2010] was implemented between the regression Pearson residuals and the covariate Driver1 (see Table 3.9). The tau coefficient value of -0.194 concludes to a low negative correlation. Finally, the dependence between the individual effects μ_i and the regressor variable was assessed by the means of Kendall rank correlation [Chok, 2010]. The output coefficient is presented in Talble 3.10 and is higher than 50%, suggesting a rather strong correlation among them. This result was expected after the rejection of the Random effects model from the Wald test in Table 3.6.

Table 3.8: LM test for heteroscedasticity in fixed effect logit model

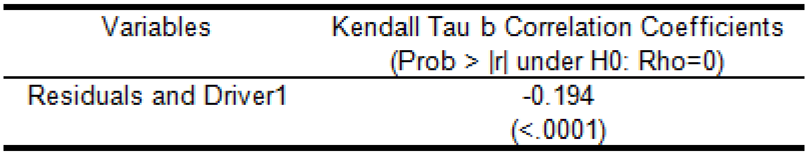


H0: $\sigma_i^2 = \sigma^2$ for all i

LM = 3.36

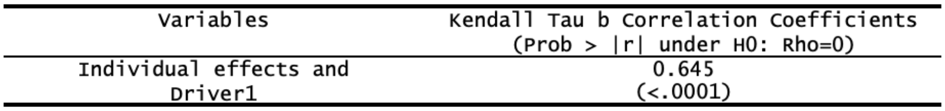
Prob > $\chi^2(1)$ = 0.0578

Table 3.9: Correlation analysis among regression Pearson residuals and the explanatory variable for fixed effect logistic model



Variables	Kendall Tau b Correlation Coefficients (Prob > |r| under H0: Rho=0)
Residuals and Driver1	-0.194 (<.0001)

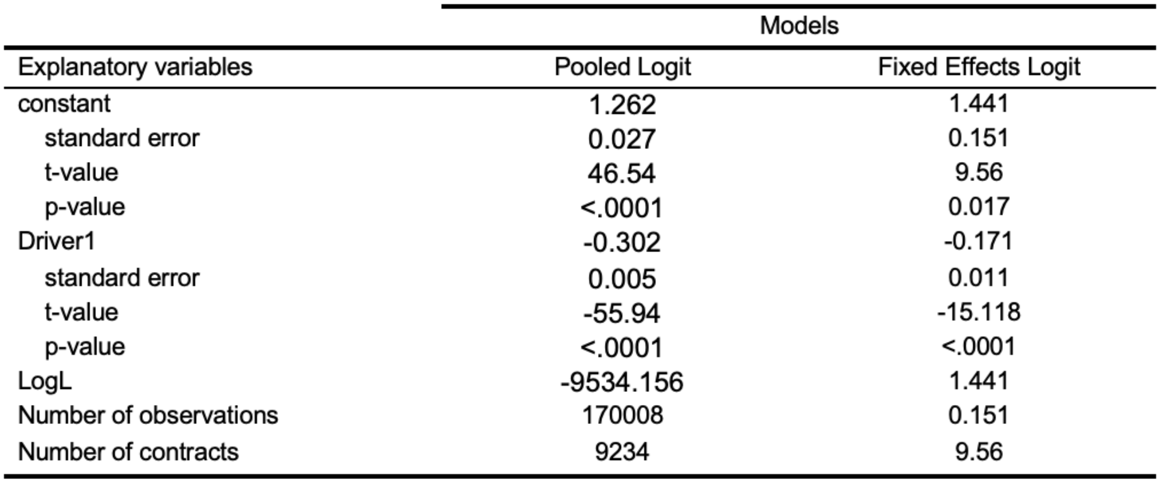
Table 3.10: Correlation analysis among regression individual fixed effects and explanatory variables



Variables	Kendall Tau b Correlation Coefficients (Prob > |r| under H0: Rho=0)
Individual effects and Driver1	0.645 (<.0001)

In conclusion, the two models which will be compared, Pooled logit and Fixed effect logit models, for modelling *Cure rate* regarding the defaulted units in a mortgage portfolio are shown in Table 3.11 below.

Table 3.11: Regression analysis results for Pooled Logit model and Fixed effects bias reduction Logit model for the defaulted Cure rate model



	Models	
Explanatory variables	Pooled Logit	Fixed Effects Logit
constant	1.262	1.441
standard error	0.027	0.151
t-value	46.54	9.56
p-value	<.0001	0.017
Driver1	-0.302	-0.171
standard error	0.005	0.011
t-value	-55.94	-15.118
p-value	<.0001	<.0001
LogL	-9534.156	1.441
Number of observations	170008	0.151
Number of contracts	9234	9.56

Using equations 3.1.1 and 3.1.2 the two estimated final models are given by the following Pooled logit model:

$$Cure\ rate_{it} = \frac{1}{1 + \exp -\left(1.262 - 0.302 \cdot Driver1_{it}\right)}, \tag{3.1.4}$$

and the following Fixed effect logit model:

$$Cure\ rate_{it} = \frac{1}{1 + \exp -\left(1.441 + \mu_i - 0.171 \cdot Driver1_{it}\right)}. \tag{3.1.5}$$

3.1.2 Model performance

After constructing a model, it is very important to assess the accuracy of its predictions. According to [Giancristofaro and Salmaso, 2007], when we evaluate the predictive ability of a model it is not enough to assess its performance only on the data used to fit it, because the results will be biased in favour of the model. Therefore, it is better to also evaluate its predictive power on an independent sample of the same population. When a model performs better for the data used to develop it than for any other sample, then its resulting estimates are sample-specific without generalizability. As [Picard and Cook, 1984] mentioned in their paper, this phenomenon is often called *principle of optimism*. For this reason, a backtesting procedure was applied to the two *Cure rate* models using In-sample and Out-of-sample tests, testing the predictive power of the models on the units used to develop them and also on excluded units from the development dataset. The most important questions rely on whether the models are able to distinguish correctly the cured from non-cured units, and how accurate the predicted rates to the observed are [Giancristofaro and Salmaso, 2007].

2.1.2.a In-sample test

In this test, the observations which will be used to validate the models are part of the construction datasets, explaining the name “In-sample” test. Specifically, the development dataset is the entire dataset. All units that were observed in the dataset to be in default during the last year were then used as a validation dataset. Therefore, the models 3.1.4 and 3.1.5 were applied on the validation data, and, a number of statistical tests were employed to evaluate their discriminatory and calibration ability.

Discriminatory power

According to [BCBS, 2006, Chap 3] “The procedure of applying a classification tool to an obligor for an assessment of her or his future status is commonly called discrimination.” Therefore, discriminatory power of a model is the ability to discriminate correctly among different status, such as defaulting and non-defaulting clients. Our models will be assessed regarding their ability to distinguish correctly the cured loans from the non-cured.

Model accuracy. In order to assess the ability of the model to adequately discriminate the unit, one translates the occurence of a cure event into a binary classification problem. To this matter, one artificially converts a continuous value, here the probability of curing, into a binary variable by defining a cutoff value c, with test results being designated as positive or negative depending on whether the resultant value is higher or lower than the defined cutoff. More formally,

$$\hat{y} = \mathbf{1}_{\{T \geqslant c\}}, \tag{3.1.6}$$

where T denotes the predicted probability of cure and c a threshold constant. For the discriminatory power of the models the sensitivity (Se) and specificity (Sp) measures were calculated as well as the false cure predictions (False Positive fraction / FPF) and false non-cure prediction (False Negative fraction / FNF) [López-Ratón et al., 2014]. Se(c) is the probability that the model correctly predict that a counterparty will cure, given a specified threshold probability c, whereas Sp(c) denotes the probability that a model correctly classifies a non-cured counterparty as non-cured, given the threshold c. The FPF and FNF are the probabilities for wrongly status classification by the model. FPF occurs if the model incorrectly predict that a loan will cure and FNF would occur in case a loan was classified mistakenly as non-cured. These measures are defined in mathematical terms as:

$$\begin{aligned}
Se(c) &= \mathbb{P}(\hat{y} = 1 | y = 1), && (3.1.7)\\
Sp(c) &= \mathbb{P}(\hat{y} = 0 | y = 0), && (3.1.8)\\
FPF &= \mathbb{P}(\hat{y} = 1 | y = 0), && (3.1.9)\\
FNF &= \mathbb{P}(\hat{y} = 0 | y = 1), && (3.1.10)
\end{aligned}$$

where y denotes the observed cure variable [López-Ratón et al., 2014].

According to [Greene, 2012], one must choose wisely the value of threshold c and not randomly. For example, by selecting $c = 0.5$ in a dataset with very small amount of zeros or ones regarding the response variable y the prediction rule (5) might not be able to predict any zero (or one). For this reason, four different methods were considered to find an optimal cutpoint c[López-Ratón et al., 2014]. First, the Youden Index method was used where the cutpoint c maximizing this quantity $YI(c) = Se(c) + Sp(c) - 1$ [Youden, 1950], whereas in the ROC01 method c minimizes the distance among the ROC plot and point $(0, 1)$, i.e minimizes this measure $(Sp(c) - 1)^2 + (Se(c) - 1)^2$ [Metz, 1978]. The next method which minimizes the $|Sp(c) - Se(c)|$ in order to achieve as close to exactly the same as possible percentages of correct cure and non-cure classification is called SpEqualSe. In the last approach, $MaxSpSe$, c maximizing the $min(Sp(c), Se(c))$. For more details about these approaches and alternative ones see [López-Ratón et al., 2014].

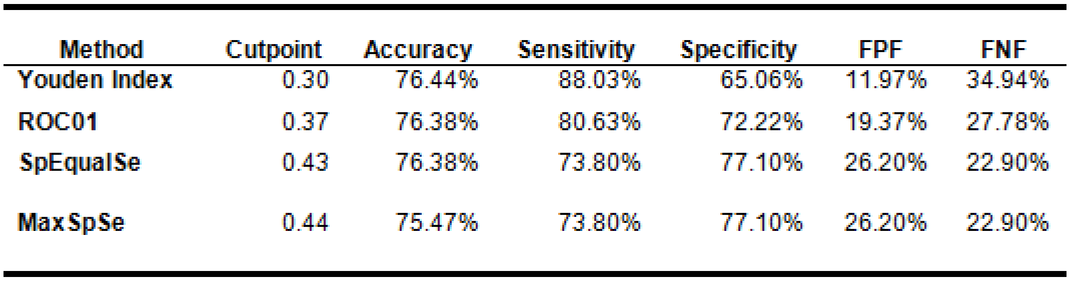
For the four aforementioned methods the four different measures and an overall accuracy of the two models are computed and shown in Tables 3.12 and 3.13 below. The accuracy , or "the fraction of the study population that is decided correctly" [Metz, 1978] is the proportion of the correct classified contracts over all the defaulted ones. What is interesting in these tables is the surprisingly different optimal threshold probabilities c for the two models. For the FE logit model a probability close to 0.75 was proposed as the optimal one from all the four methods, whereas for the Pooled logit model the optimum value for c is different for every method and it ranges between 0.30 and 0.44. This means, that most of the times the FE model assigns probabilities equal or higher than 0.75 to the cured loans and at the same time the other model assigns much lower probabilities to the cured contracts, as is reflected from its low optimal cutpoints.

Furthermore, the results in Tables 3.12 and 3.13 indicate very high values of sensitivity and specificity for the Fixed effect (FE) model under all the four approaches. Pooled logit model scored lower number of accurate predictions than the FE model regarding the four different optimal thresholds. All in all, the FE model managed to discriminate almost all the contracts correctly under the threshold $c = 0.74$, reaching overall accuracy very close to 100%. While the Pooled logit model predicts wrongly cure for non-cured contracts in a level of 34.94% and non-cure for cured contracts 11.97%, based on Youden Index method, scoring an accuracy 76.44%. In the case of MaxSpSe approach and SpEqualSe where the cutpoint c is higher, the pooled model has about the same accuracy as in Youden Index method and the FPF is reduced at 22.9%, but the FNF grows at 26.20%. This phenomenon is mentioned in [Greene, 2012], where he says that by changing the value of the threshold probability c so as to classify correctly more observations with $y = 1$, results to increase the incorrectly classifications for data with $y = 0$.

In Appendix B plots of the $Se(c)$ and $Sp(c)$ measures for the two models against nine different cutpoint probabilities can be found. From these plots it can be seen that the FE model can predict actual cure contracts better than the pooled model for all the thresholds. Specifically, when $c \in \{0.1, \ldots, 0.7\}$ the model with the unit-specific effects predicts successfully all the cured loans. However, as regards the non-cure contracts the FE model predict actual non-cured contracts more accurate for $c \geqslant 0.3$.

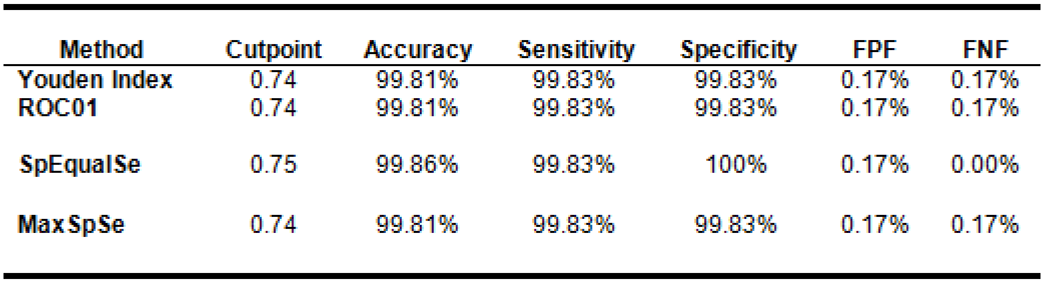
To conclude, according to the aforementioned statistics measures the Pooled model has less discriminatory power than the FE model for thresholds $c \in \{0.3, \dots, 0.9\}$, but when $c = 0.1$ and $c = 0.2$ its specificity is better than the one of FE model.

Table 3.12: Model accuracy results for the Pooled logit model



Method	Cutpoint	Accuracy	Sensitivity	Specificity	FPF	FNF
Youden Index	0.30	76.44%	88.03%	65.06%	11.97%	34.94%
ROC01	0.37	76.38%	80.63%	72.22%	19.37%	27.78%
SpEqualSe	0.43	76.38%	73.80%	77.10%	26.20%	22.90%
MaxSpSe	0.44	75.47%	73.80%	77.10%	26.20%	22.90%

Table 3.13: Model accuracy results for the Fixed Effects logit model

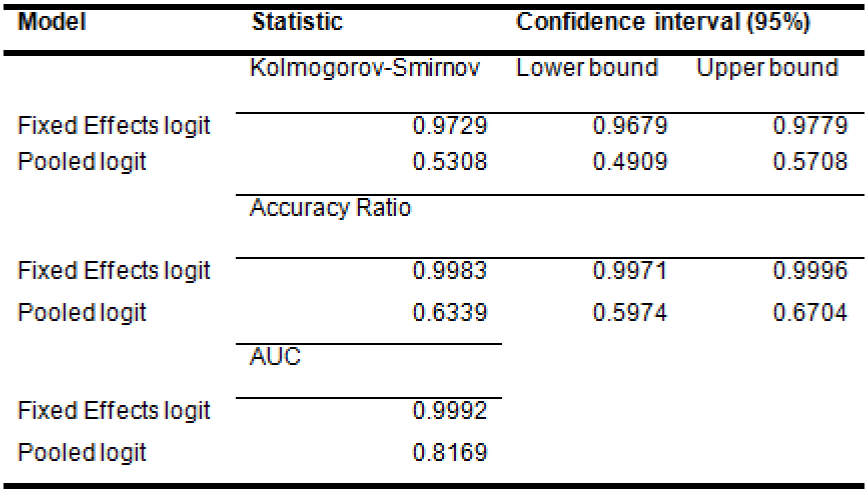


Method	Cutpoint	Accuracy	Sensitivity	Specificity	FPF	FNF
Youden Index	0.74	99.81%	99.83%	99.83%	0.17%	0.17%
ROC01	0.74	99.81%	99.83%	99.83%	0.17%	0.17%
SpEqualSe	0.75	99.86%	99.83%	100%	0.17%	0.00%
MaxSpSe	0.74	99.81%	99.83%	99.83%	0.17%	0.17%

Kolmogorov-Smirnov test, Accuracy Ratio and AUC. The discriminatory ability of the two models was also assessed by computing the Kolmogorov-Smirnov (KS) statistic, Accuracy Ratio (AR), the Area Under the ROC Curve (AUC). The KS test can be used to examine whether the two data samples originate from the same distribution, with test statistic the maximum distance of the two cumulative distributions functions (CDF) [Řezáč and Řezáč, 2011, Rasero, 2006]. In our case, the two data samples are the cure and non-cured loans from the defaulted loans in-sample validation dataset. Hence, the higher the value of this test statistic the higher the distance among the two CDF, and therefore, the better discrimination among the cured and non-cured loans.

Furthermore, the AR is a summary index of the Cumulative Accuracy Profiles (CAP) curve, which is based on the cumulative probabilities of cured and non-cured loans for the entire dataset. Thus, AR measures the model's predictive accuracy across all the data [BCBS, 2006, Chap 3]. Moreover, The AUC statistic is the area under the Receiver Operating Characteristics (ROC) curve, which is the plot of Se(c) against 1-Sp(c)

Table 3.14: Kolmogorov-Smirnov test statistics, Accuracy Ratios, Area under the ROC curves and confidence intervals for the in-sample validation dataset



Model	Statistic	Confidence interval (95%)	
	Kolmogorov-Smirnov	Lower bound	Upper bound
Fixed Effects logit	0.9729	0.9679	0.9779
Pooled logit	0.5308	0.4909	0.5708
	Accuracy Ratio		
Fixed Effects logit	0.9983	0.9971	0.9996
Pooled logit	0.6339	0.5974	0.6704
	AUC		
Fixed Effects logit	0.9992		
Pooled logit	0.8169		

across different probability values of c [Engelmann et al., 2003, Řezáč and Řezáč, 2011]. This statistic gives the probability that a randomly selected cured loan has assigned with higher cure likelihood than a randomly selected non-cured loan. Additionally, AUC is linearly related with the AR and therefore they use the same information and lead to the same conclusions [Řezáč and Řezáč, 2011, Medema et al., 2009]. For proof see [Rasero, 2006]. The AUC statistic is also known as c-statistics or AUROC [Řezáč and Řezáč, 2011]. Regarding the statistics AR and AUC, the higher their output value the better the model distinguishes between cured and non-cured loans [Medema et al., 2009].

The resulted test statistics are presented in Table 3.14. The high ability of the FE model to differentiate the contracts among cured and non-cured, is deduced from the very high statistics output values shown in the table. Specifically, all three statistics score almost 1.00 for the FE model. Whilst, the Pooled model scores are lower than the FE model output values. Consequently, by comparing the results of the two models, we can conclude that the model which incorporates loan-specific fixed effects outerperform the pooled model under these tests and it can discriminate better the loans.

Calibration quality analysis

According to [Giancristofaro and Salmaso, 2007] the calibration is " a measure of how close the predicted probabilities are to the observed rate of the positive outcome for any given configuration of the independent variables of the model". In addition, [on Banking Supervision, 2004] states "Banks must regularly compare realized default rates with estimated PDs for each grade and be able to demonstrate that the realized default rates are within the expected range for that grade. Banks using the advanced IRB approach must complete such analysis for their estimates of LGDs and EADs."

Hence, with regard the calibration quality of a model the differences of the predicted probabilities that a defaulted client will cure and the realised cure rates, must be studied.

Binomial test The calibration ability of the two models was firstly assessed by means of a two-sided Binomial test. This test can be used to examine whether the observed number of cured rates within different classes is consistent with the predicted cure probabilities (BCBS Chapter 3, 2005b). The null hypothesis of this test is that the estimated cure rate of the class is correct and is normally distributed. A model can be considered as "Conservative" when the upper bound of its prediction's confidence interval is lower than the realised probability, while when the lower bound of the confidence interval is higher than the realised cure ratio the model is classified as "Optimistic". If none of the above cases occur, we have an "Accurate" model and the null hypothesis is not rejected.

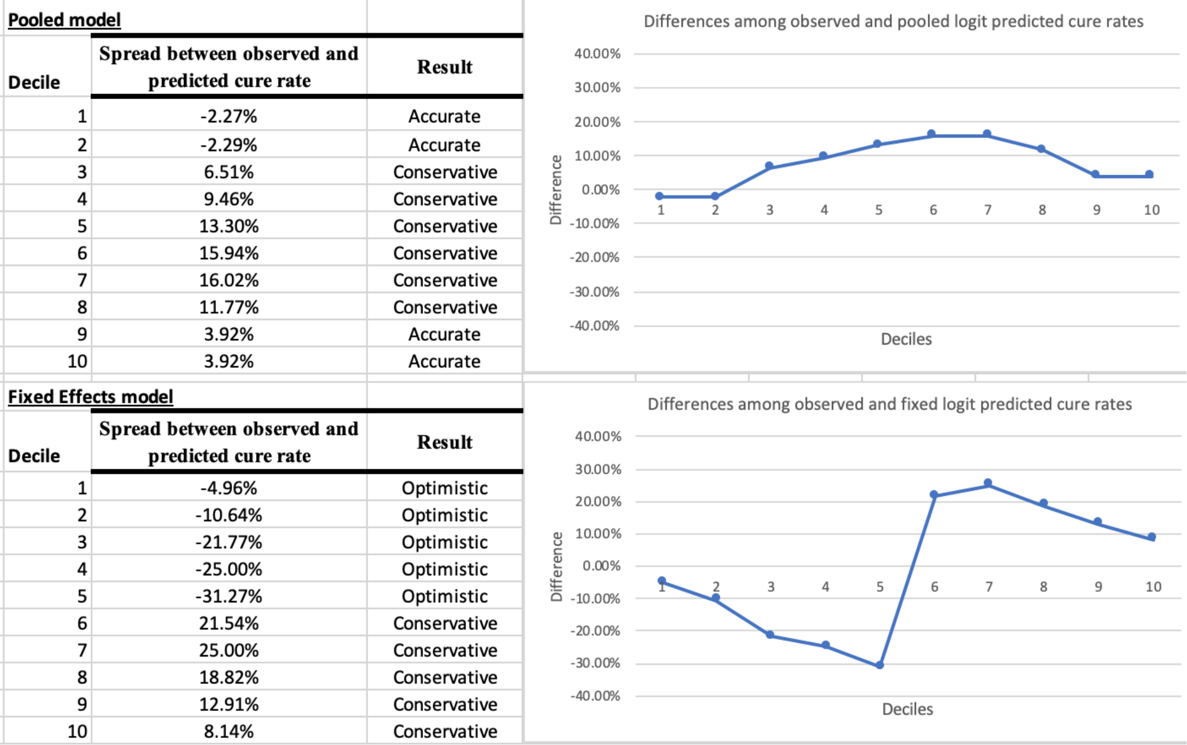
The Tables 3.15 and 3.16 below present the results of the assessment regarding the predictive accuracy of the two models in the in-sample period. For the binomial test the validation dataset was divided into 10 groups of the same size. The test was employed on two levels, on the 10 deciles as well as on the entire one-year dataset. Tables 3.15 and 3.16 show the differences between the realized and the observed cure rates which were calculated for the entire validation dataset and for every decile, respectively. When the difference is negative, it means that the predicted cure rate is higher than the observed one and when it is positive means the opposite. The last column of the tables indicates if the estimated cure rates are "Conservative", "Optimistic" or "Accurate" based on the aforementioned definitions. Therefore, when the difference between the cure rates is negative and the lower bound of the predicted cure rate confidence interval is higher than the realised cure ratio the model is classified as "Optimistic". On the opposite situation, where the difference is positive and the upper bound of its prediction's confidence interval is lower than the realised probability, the estimations are called "Conservative". The exact observed and predicted cure rates are confidential information and for this reason are not shown in this thesis. For a better understanding of the different predictions that the two models give, the difference between the observed and predicted cure rates are illustrated graphically in Table 3.16.

What is striking in the results, is the performance of the FE model which predicts optimists cure rates for the half deciles and conservative for the rest of them. For none of these classes managed to estimate an accurate probability. Only when the entire validation dataset is considered as one class, in Table 3.15, succeed to predict a cure rate almost equal to the realized one. In contrast, the Pooled model has an overall "Conservative" character for the entire dataset which align with its predictions for most of the classes. This means that the model predicts a lower number of defaulted loans that will cure than the real one. But, for four classes managed to estimate accurate rates. It is better for a company to predict that less clients will cure and calculate higher loss than the realized, in other words to get conservative results, rather than expecting a higher cure rate than the true one and be in the unpleasant position to find out that its loss will be higher than the expected one. The last case will happen with optimistic results. Consequently, if we considered only the results for the different deciles, the bank's model calibrates better the data since gives accurate and conservative results whereas the FE model either predicts higher or lower rates than the observed.

Table 3.15: Binomial test results for the entire in-sample validation dataset

Model	Spread between observed and predicted cure rate	Result
Pooled logit	7.63%	Conservative
Fixed Effect logit	-0.72%	Accurate

Table 3.16: Binomial test results for the in-sample validation dataset per decile



Pooled model

Decile	Spread between observed and predicted cure rate	Result
1	-2.27%	Accurate
2	-2.29%	Accurate
3	6.51%	Conservative
4	9.46%	Conservative
5	13.30%	Conservative
6	15.94%	Conservative
7	16.02%	Conservative
8	11.77%	Conservative
9	3.92%	Accurate
10	3.92%	Accurate

Differences among observed and pooled logit predicted cure rates

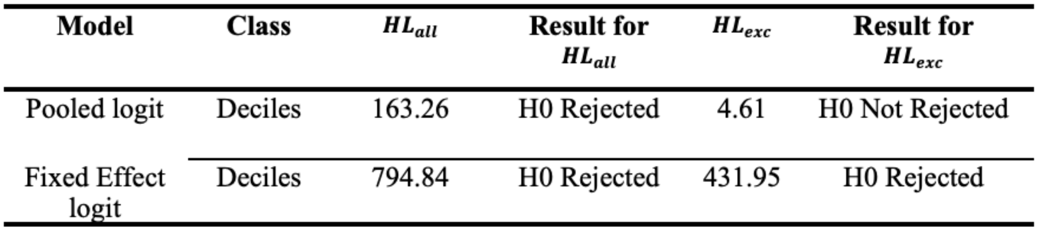
Fixed Effects model

Decile	Spread between observed and predicted cure rate	Result
1	-4.96%	Optimistic
2	-10.64%	Optimistic
3	-21.77%	Optimistic
4	-25.00%	Optimistic
5	-31.27%	Optimistic
6	21.54%	Conservative
7	25.00%	Conservative
8	18.82%	Conservative
9	12.91%	Conservative
10	8.14%	Conservative

Differences among observed and fixed logit predicted cure rates

Hosmer-Lemeshow test The calibration power of the two models was also tested by means of one-sided Hosmer-Lemeshow (HL) test which tests the results of several classes simultaneously [Giancristofaro and Salmaso, 2007]. The HL test results in one outcome covering all rating classes and it examines whether the predicted cure probabilities assigned to the different classes, which used in the binomial test before, are not significantly different from the observed cure rates. The test statistic follows a Chi-square distribution and the null hypothesis of all predicted cure rates being correct is rejected when the test statistic is larger than the 95%-quantile of the distribution.

Table 3.17: Hosmer-Lemeshow test results for the out-of-sample validation dataset per decile and year



Model	Class	HL_{all}	Result for HL_{all}	HL_{exc}	Result for HL_{exc}
Pooled logit	Deciles	163.26	H0 Rejected	4.61	H0 Not Rejected
Fixed Effect logit	Deciles	794.84	H0 Rejected	431.95	H0 Rejected

As we mentioned earlier the conservative estimations are not harmful for the bank, in contrast with the optimistic ones. Thus, the HL test statistic was computed first based on all the predictions per class (HL_{all}), and thereafter excluding the conservative cases (HL_{exc}).

From the Table 3.17 we see that by considering all the estimations that are not classified as accurate to be "bad" predictions, then both models are failed to predict correct cure rates, since there test statistics HL_{all} are higher than the chi-square critical value. However, when the conservative estimations are excluded from the calculations the Pooled model's accuracy changed and the null hypothesis was not rejected. This comes in contradiction with the resulting test statistic regarding the FE, which remains very high. These results were expected given the many conservative outputs of the Pooled model and optimistic of FE model, as regards the binomial test. Therefore, under the condition that the conservative estimations are not "bad" estimations, the Pooled model performed better than the FE model with respect to their calibration ability.

2.1.2.b Out-of-sample test.

On the contrary to the in-sample method, in an out-of-sample test, the model performance is validated on a dataset out of the training dataset. Given [Giancristofaro and Salmaso, 2007] an accredited method to perform out-of-sample test is by $Data - splitting$, where the entire dataset is splitted randomly into two sub-datasets, obtaining the development and testing datasets. The split ratio, most of the times, is between two thirds and three quarters, with the biggest subsample used for the training dataset and the remaining set as the validation dataset. For other methods to obtain the development and validation dataset consult [Giancristofaro and Salmaso, 2007]. However, a fixed effects panel data model can be applied only on data with the same individuals that are used to build the model and not on new ones. The reason for this is the individual-specific fixed effect which is unique for every individual. As a result, the development and validation datasets will be different but dependent. In this thesis that data are unbalanced and a simple data split into two subdatasets is not appropriate. Therefore, all defaulted units between the first 3 years of the 5-years dataset were chosen to construct the model, and from the units with remaining years were selected only the ones which were observed in the development dataset. For that reason, the estimation of new models was nessecary. The resulting models based on the contracts which were get into default or were already defaulted between the fisrt 3 years, are:

Pooled logit model:

$$Cure\ rate_{it} = \frac{1}{1 + \exp -\big(1.176 - 0.311 \cdot Driver1_{it}\big)}, \tag{3.1.11}$$

and Fixed effect logit model:

$$Cure\ rate_{it} = \frac{1}{1 + \exp -\big(1.661 + \mu_i - 0.091 \cdot Driver1_{it}\big)}. \tag{3.1.12}$$

where μ_i measures the specific fixed effects for each contract i throughout the years t.

The two models were assessed regarding their discriminatory and calibration power, as in the in-sample test.

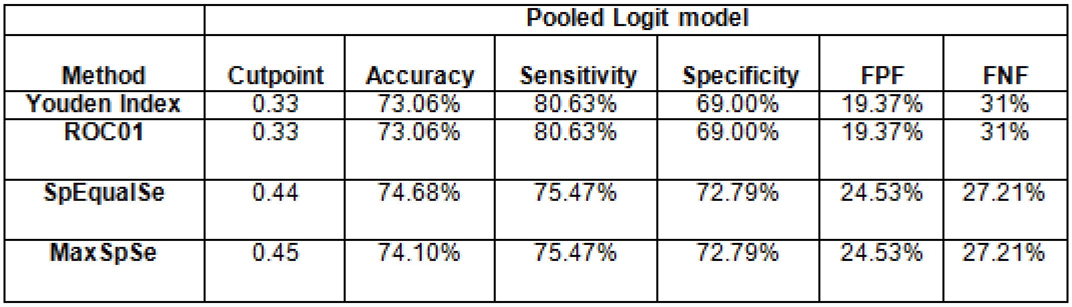
Discriminatory power

Model accuracy. As in the in-sample backtesting test, the accuracy, sensitivity $Se(c)$, specificity $Sp(c)$, FPF and FNF measures were calculated for the two models [Metz, 1978, Steyerberg et al., 2010, López-Ratón et al., 2014]. Four different approaches were used to get an optimal cutpoint c for the prediction rule in 3.1.6 [López-Ratón et al., 2014]. The resulting threshold probabilities c are again very different across the two models indicating the different probabilities for cure that the models assign to the same contracts. The Table 3.18 depicts that for the FE model the optimum c is approximately 0.70 in all cases and using this it correctly classifies the 90.44% until 96.48% of cured units. Fortunately, these high sensitivity values are not harmful for the specificity measure, which lies between 87.69% and 90.45%. At the same time, the Pooled model with cutpoint value 0.33 manages to capture 80.63% of the cured contracts and 69.00% of the non-cured ones. Using the SpEqualSe and MaxSpSe methods the specificity rose slightly to 72.79% as the cutpoint value increased from 0.33 to 0.44 and 0.45, respectively, but the sensitivity yields a reduction by 5%.

In addition, $Se(c)$ and $Sp(c)$ measures were computed for a number of different threshold probabilities c for the two models. Plots in Appendix B? show graphically that the FE model always performs better when it comes to classify correctly a cured loan. Specifically, for $c \in \{0.1, \dots, 0.5\}$ the FE model predicts successfully all the actual cured loans. As regards the false non-cured predictions (FNF) the Pooled model suffers from a lower discriminatory power than the other model for all the values for c except when c=0.1. In that case the Pooled model manages to classify correctly 44.38% of the non-cured loans, whereas the FE model captures only the 16.94%.

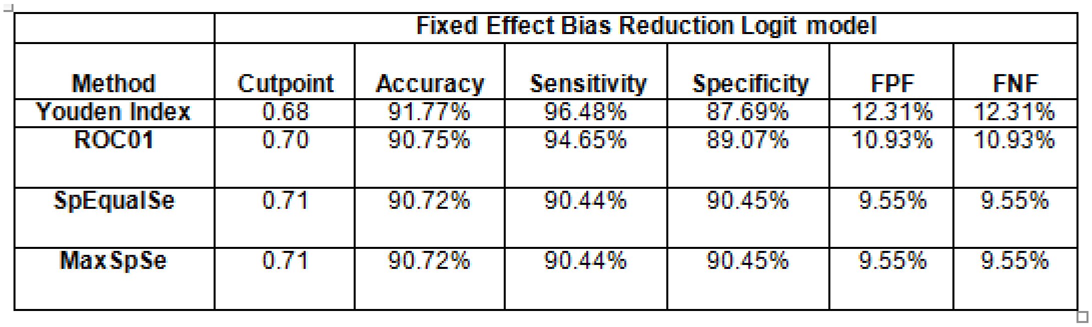
Overall, it is noticeable from Table 3.18 and plots in Appendix B that the FE model is better in distinguishing cured from non-cured loans in the out-of-sample validation dataset than the pooled model. This is evidenced by the higher accuracy, $Sp(c)$ and $Se(c)$ values of the first model over the pooled one. Only when $c = 0.1$ the model which ignores the loan-specific effects performs better regarding the specificity measure.

Table 3.18: Model accuracy results for the Pooled logit model



	Pooled Logit model					
Method	Cutpoint	Accuracy	Sensitivity	Specificity	FPF	FNF
Youden Index	0.33	73.06%	80.63%	69.00%	19.37%	31%
ROC01	0.33	73.06%	80.63%	69.00%	19.37%	31%
SpEqualSe	0.44	74.68%	75.47%	72.79%	24.53%	27.21%
MaxSpSe	0.45	74.10%	75.47%	72.79%	24.53%	27.21%

Table 3.19: Model accuracy results for the Fixed Effects logit model



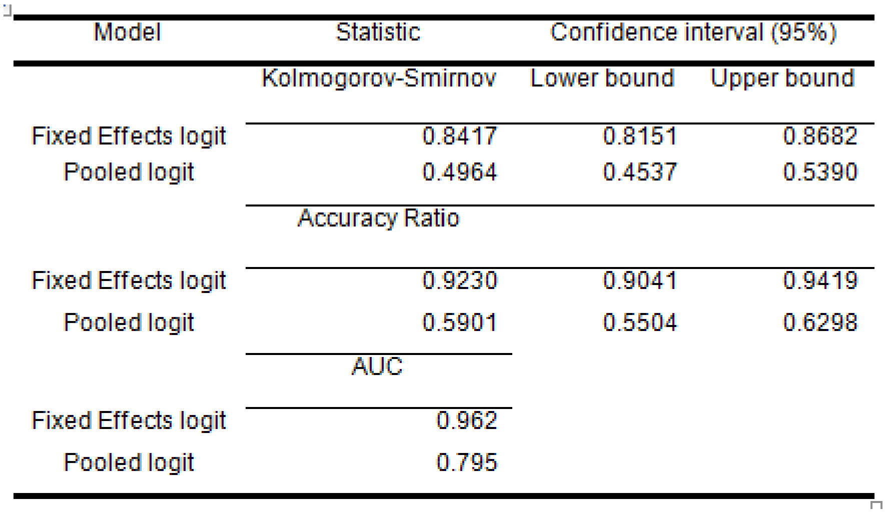
	Fixed Effect Bias Reduction Logit model					
Method	Cutpoint	Accuracy	Sensitivity	Specificity	FPF	FNF
Youden Index	0.68	91.77%	96.48%	87.69%	12.31%	12.31%
ROC01	0.70	90.75%	94.65%	89.07%	10.93%	10.93%
SpEqualSe	0.71	90.72%	90.44%	90.45%	9.55%	9.55%
MaxSpSe	0.71	90.72%	90.44%	90.45%	9.55%	9.55%

Kolmogorov-Smirnov test, Accuracy Ratio and AUC For the forecasting performance of the two models, three more statistics were calculated. The AR , KS and AUC statistic values [Engelmann et al., 2003, Řezáč and Řezáč, 2011](BCBS Ch.3, 2005) are presented in Table 3.20. The FE model has KS statistic 0.84 and AR 0.92, meaning that it achieves to make the distinction perfectly among cured and non-cured contracts. On the contrary, the Pooled logit reaches KS test statistic value 0.49 and AR at the level of 0.59. Therefore, it is considered to be not good as the FE model based on their discriminatory power. Another statistic that captures the different performance of the two models is the area under the ROC curve (AUC) [Engelmann et al., 2003, Řezáč and Řezáč, 2011]. The model which takes into consideration the loan-specific fixed effect appears to score any random client from the dataset who is going to cure with a higher event probability than any non-cure client, since its AUC value is 0.962. The bank's model accomplishes that in 79.5% of the cases. All in all, all three statistics indicate that the FE logit model performs better regarding discrimination of the cure/non-cure clients than the pooled logit model.

Calibration quality analysis

Binomial test The binomial test (BCBS Chapter 3, 2005b) was carried out on the validation dataset in the out-of-sample period. First, the test was applied on the entire validation dataset, and then on every year of the dataset separately (Table 3.21). Fol-

Table 3.20: Kolmogorov-Smirnov test statistics, Accuracy Ratios, Area under the ROC curves and confidence intervals for the out-of-sample validation dataset

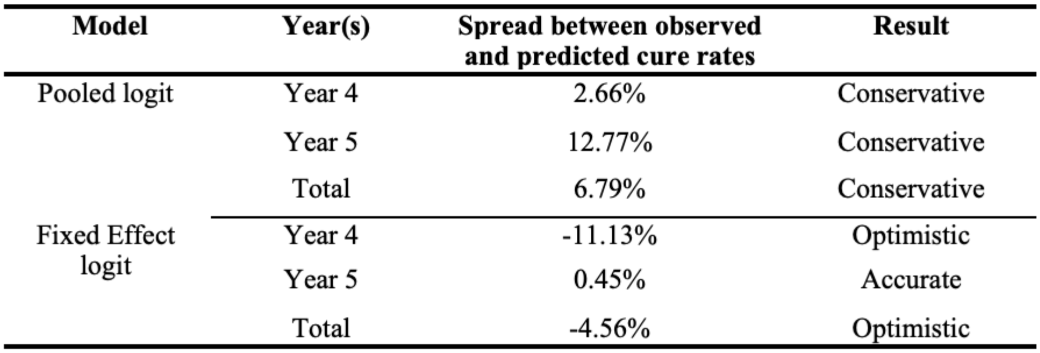


Model	Statistic	Confidence interval (95%)	
	Kolmogorov-Smirnov	Lower bound	Upper bound
Fixed Effects logit	0.8417	0.8151	0.8682
Pooled logit	0.4964	0.4537	0.5390
	Accuracy Ratio		
Fixed Effects logit	0.9230	0.9041	0.9419
Pooled logit	0.5901	0.5504	0.6298
	AUC		
Fixed Effects logit	0.962		
Pooled logit	0.795		

lowing that, the data were grouped into deciles and predicted cure rates were generated for every decile for both models (Table 3.22). In Tables 3.21 and 3.22 the differences between the realized and the observed cure rates are shown. Moreover, in the last column of the tables the model is classified as "Conservative", "Optimistic" or "Accurate", based on the distance between the observed and realized rates. For more details about these terms see the Binomial test in the In-sample test section.

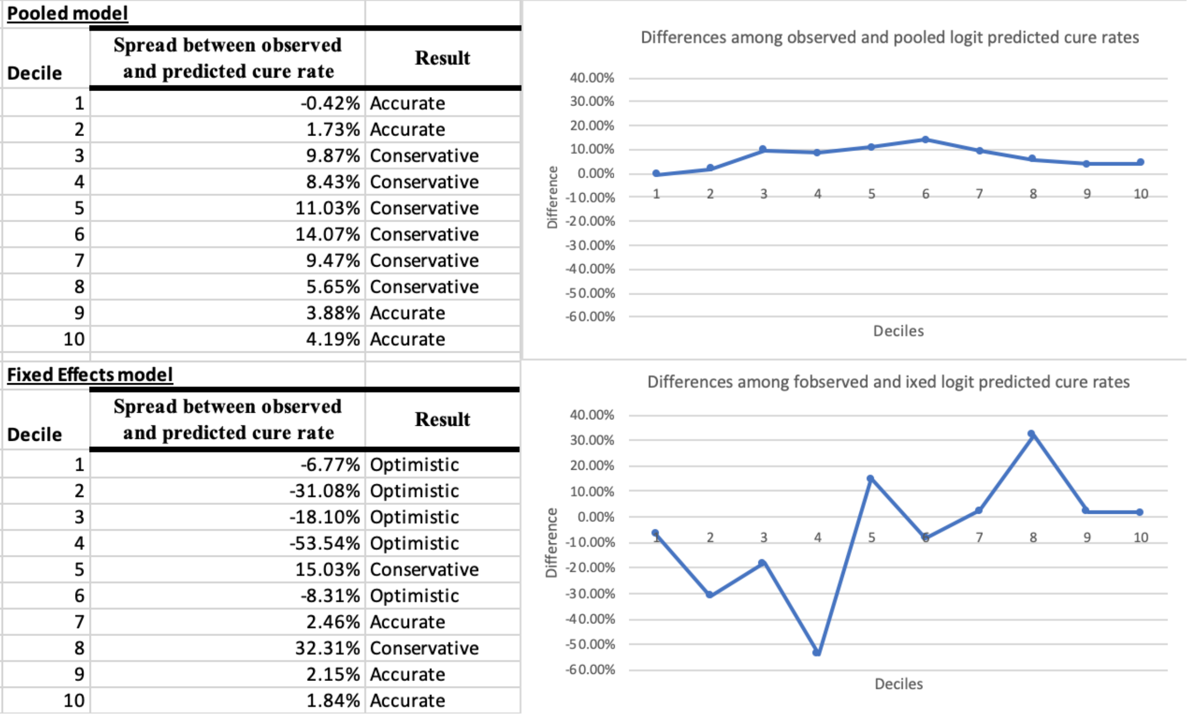
The results in Tables 3.21 and 3.22 regarding the Pooled model are the same as in the in-sample test. The bank's model estimates lower number of cured loans than the observed when the test employed on the entire dataset and on each year separately (Table 3.21) and on the most of deciles (Table 3.22). In the rest of deciles, the Pooled model predicts probabilities very close to the true ones which are ranked as accurate. While, the FE model has an overall optimistic character, despite some accurate and conservative estimations. This is caused because most of the times the FE model predicts higher cure probabilities than the realised ones. Additionally, from the first graph in Table 3.22 we can see that the Pooled model always predict rates that are higher than the realized ones with maximum error at 14.07% level. On the other hand, the graph which illustrates the differences among fixed effect predicted and observed rates, shows that the FE model gives higher errors compared to the pooled logit errors. Therefore, considering that for a bank is better to be “Conservative” rather than “Optimistic”, one can result that the Pooled model is better as regards the calibration power.

Table 3.21: Binomial test results for the entire out-of-sample validation dataset



Model	Year(s)	Spread between observed and predicted cure rates	Result
Pooled logit	Year 4	2.66%	Conservative
	Year 5	12.77%	Conservative
	Total	6.79%	Conservative
Fixed Effect logit	Year 4	-11.13%	Optimistic
	Year 5	0.45%	Accurate
	Total	-4.56%	Optimistic

Table 3.22: Binomial test results for the out-of-sample validation dataset per decile



Pooled model

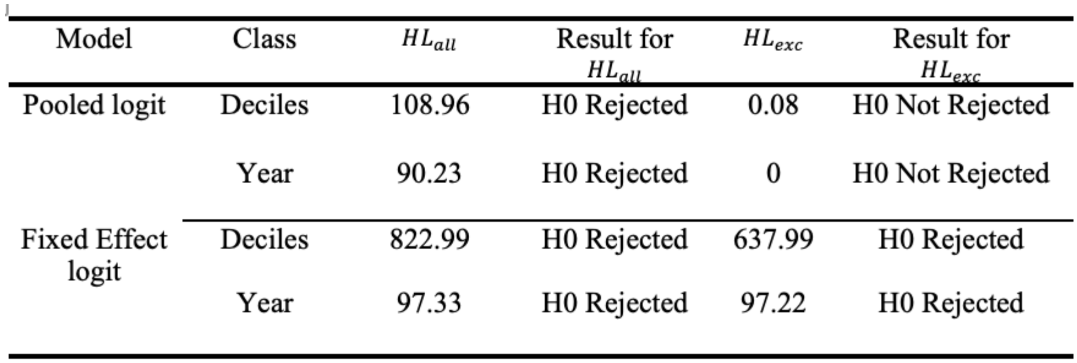
Decile	Spread between observed and predicted cure rate	Result
1	-0.42%	Accurate
2	1.73%	Accurate
3	9.87%	Conservative
4	8.43%	Conservative
5	11.03%	Conservative
6	14.07%	Conservative
7	9.47%	Conservative
8	5.65%	Conservative
9	3.88%	Accurate
10	4.19%	Accurate

Fixed Effects model

Decile	Spread between observed and predicted cure rate	Result
1	-6.77%	Optimistic
2	-31.08%	Optimistic
3	-18.10%	Optimistic
4	-53.54%	Optimistic
5	15.03%	Conservative
6	-8.31%	Optimistic
7	2.46%	Accurate
8	32.31%	Conservative
9	2.15%	Accurate
10	1.84%	Accurate

Hosmer-Lemeshow test. Moreover, the Hosmer-Lemeshow test [Giancristofaro and Salmaso, 2007] was applied on the out-of-sample validation dataset which was grouped first by deciles and then by ultimo year. It tests whether the predicted cure probabilities assigned to the different classes, are not significantly different from the observed cure rates. Therefore, the null hypothesis of this test is that all predicted cure rates are correct. The HL test statistic was computed first based on all the predictions per class (HL_all) and thereafter excluding the conservative cases (HL_exc), as we did in the In-sample test. The results in Table 3.23 reflect the same conclusions as in the in-sample method. In more details, the Pooled model achieves to estimate accurate probabilities when the conservative estimations are excluded from the calculations, whereas it fails to predict accurate when all the data are taken into consideration. The FE model, on the other hand, rejects the null hypothesis in all the cases. As a consequence, if a company classifies the conservative estimations as "good" then the Pooled model is preferable than the FE.

Table 3.23: Hosmer-Lemeshow test results for the out-of-sample validation dataset per decile and year



Model	Class	HL_{all}	Result for HL_{all}	HL_{exc}	Result for HL_{exc}
Pooled logit	Deciles	108.96	H0 Rejected	0.08	H0 Not Rejected
	Year	90.23	H0 Rejected	0	H0 Not Rejected
Fixed Effect logit	Deciles	822.99	H0 Rejected	637.99	H0 Rejected
	Year	97.33	H0 Rejected	97.22	H0 Rejected

To conclude, according to in-sample and out-of-sample back-tests the FE model can discriminate better among cured and non-cured loans than a pooled model. Thing that conflicts with the poor calibration performance of the FE model and the accurate predictions of the Pooled model. According to [Diamond, 1992] this is not a surprise, since "a prediction model cannot be both perfectly reliable and perfectly discriminatory". A predictive model with excellent discriminatory power does so sacrificing its calibration ability to give accurate estimations.

3.2 The Loss-given-no cure (LGN) model

Following the Cure rate defaulted model, the LGN defaulted model is constructed in order to complete the development of LGD model for already defaulted loans. Given the particularity of the mortgage products in the Netherlands, the LGN defaulted model is split into two sub-segments depending whether loans are NHG-guaranteed or not.

3.2.1 Model construction

For the construction of the LGN model a sub-dataset was constructed with all the non-cure defaulted contracts form the 5-years dataset. Hence, we are dealing again with an unbalanced panel dataset, but now the number of contracts N is smaller than in the entire 5-year dataset. As a first step, we test the contribution of each of the dataset variables to the explanatory power of the model. For the association between the explanatory variable and the different continuous and categorical variables, the Kruskal-Wallis test and Wilcoxon-Mann-Whitney test were performed. Following this, a Kendall tau correlation analysis was performed.Considering the results of these tests, the explanatory variables x_1, x_2 for the LGN model was chosen. The results of these tests are confidential and therefore were removed from this report. Hence, here we will call the explanatory variables as $Driver1$ and $Driver5$.

Based on pooled data and in-default contracts, a multilinear regression based on $Driver1$ and $Driver5$ is built as follows:

$$LGN_{it} = \alpha + \beta_1 \cdot Driver1 + \beta_2 \cdot Driver5 + \epsilon_{it}, \tag{3.2.1}$$

where α is the itnercept, β_1 and β_2 are regression coefficeints of independent variables $Driver1$ and $Driver5$, ϵ_{it} stands for the disturbance term of each unit i in every year t.

We are going to construct the aforementioned model and the defaulted LGN model for panel data, using the same explanatory variables and datasets, and then compare their performance. The panel data model is:

$$LGN_{it} = \alpha + \mu_i + \beta_1 \cdot Driver1 + \beta_2 \cdot Driver5 + v_{it} \tag{3.2.2}$$

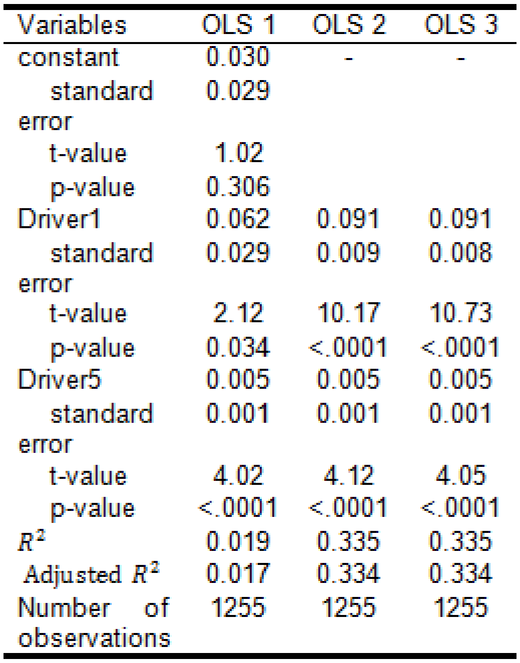
where $\mu_i + v_{i,t}$ is the error component of the model, μ_i denotes the *unobservable* effect for each unit and $v_{i,t}$ represents the remainder disturbance term. The remaining variables are the same as in pooled model.

2.2.1.i NHG portfolio

2.2.1.i.a Pool data modelling

For the NHG model, firstly, a pooled Ordinary Least Squares (OLS) regression is performed with Driver1 and Driver5 as explanatory variables in a model with LGN as dependent variable. The results are shown in Table 3.24. This model has not significant intercept. Hence, the same regression is conducted with no intercept. Driver1 and Driver5 remain both significant whereas the measure R^2 increases from 1.9% to 33.54%.

Table 3.24: Pooled OLS regression output (OLS 1), without intercept (OLS 2) and with HAC standard errors (OLS 3)

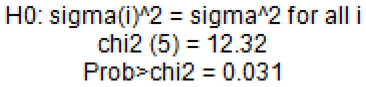


Variables	OLS 1	OLS 2	OLS 3
constant	0.030	-	-
standard error	0.029		
t-value	1.02		
p-value	0.306		
Driver1	0.062	0.091	0.091
standard error	0.029	0.009	0.008
t-value	2.12	10.17	10.73
p-value	0.034	<.0001	<.0001
Driver5	0.005	0.005	0.005
standard error	0.001	0.001	0.001
t-value	4.02	4.12	4.05
p-value	<.0001	<.0001	<.0001
R^2	0.019	0.335	0.335
Adjusted R^2	0.017	0.334	0.334
Number of observations	1255	1255	1255

Then, the model assumptions were tested. We performed both White and Durbin-Watson tests to check if the error terms suffer from heteroscedasticity and first-order autocorrelation, respectively. Details about these tests can be found in Appendix A. The null hypothesis of the White test is that the variance of the error terms is constant. This hypothesis is rejected at a significance level of 0.03 (Table 3.25). Moreover, Table 3.26 depicts the Durbin-Watson statistic and the p-values for testing positive ($Pr < DW$) and negative ($Pr > DW$) serial correlation with null hypothesis of no first-order autocorrelation. Having p-value 0.03 for positive autocorrelation and 0.97 for negative, we can conclude that our model remainder disturbances are positive autocorrelated. Therefore, heteroscedasticity-and-autocorrelation-consistent (HAC) standard errors or simply Newey–West standard errors have been estimated in order to correct the unknown form of serial correlation and heteroscedasticity in the residuals [Verbeek, 2004].

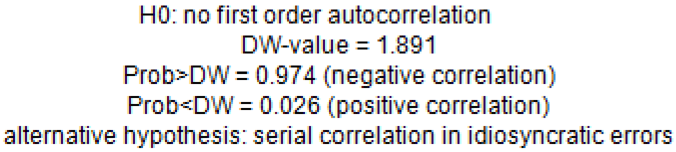
Lastly, the degree of dependence among the residuals and the two covariates was examined using the Kendall rank correlation tau ([Chok, 2010]). From the Table 3.27 we can see that the associations of Driver1 and Driver5 between the residuals are both very low and negative, since the abosolute values of the two correlation coefficients are very smaller from 0.5. The value 0.50 is considered as the critical value to determine if the correlation among two variables is high or not. Therefore, we can consider them not correlated enough to violate the assumption of independency between residuals and regressors.

Table 3.25: White test for heteroscedasticity pooled OLS model



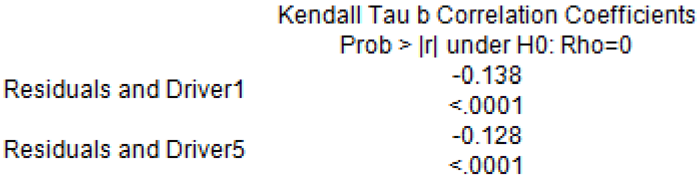
```
H0: sigma(i)^2 = sigma^2 for all i
         chi2 (5) = 12.32
        Prob>chi2 = 0.031
```

Table 3.26: Durbin-Watson test for serial correlation pooled OLS model



```
             H0: no first order autocorrelation
                     DW-value = 1.891
          Prob>DW = 0.974 (negative correlation)
           Prob<DW = 0.026 (positive correlation)
alternative hypothesis: serial correlation in idiosyncratic errors
```

Table 3.27: The correlation analysis among regression residuals and explanatory variables for the OLS pooled NHG model

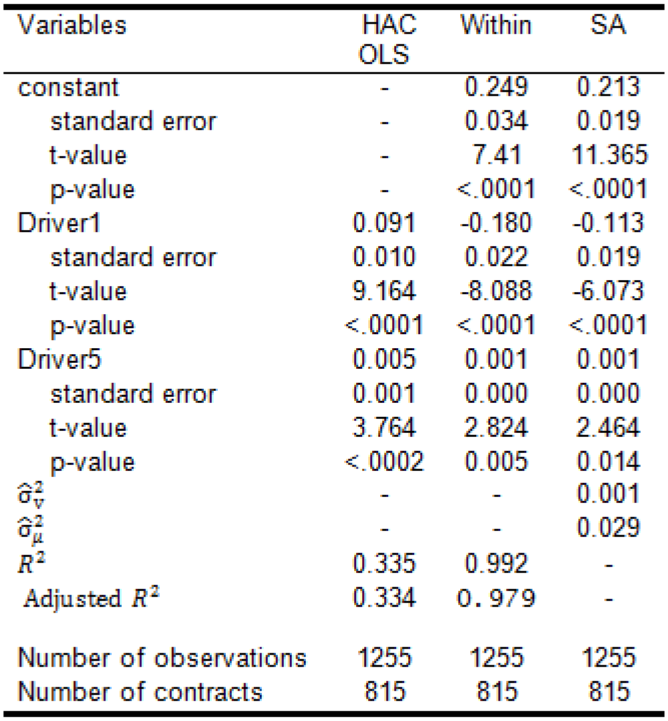


```
                          Kendall Tau b Correlation Coefficients
                              Prob > |r| under H0: Rho=0
                                        -0.138
Residuals and Driver1
                                        <.0001
                                        -0.128
Residuals and Driver5
                                        <.0001
```

2.2.1.i.b Panel data modelling

After the pooled OLS regression analysis, the relation of the explanatory variable with the independent variables has been examined through a panel data analysis with fixed and random effects. The fixed effects model tests individual differences in intercepts without making any assumption about the correlation among μ_i and the regressors $x_{i,t}$. This model is estimated by within effect estimation method. However, the random effects model assumes the individual differences to be part of the error term and independent of the $x_{i,t}$. The Swamy and Arora (SA)-type estimators captures these random effects and are illustrated in the following table (Table 3.28).

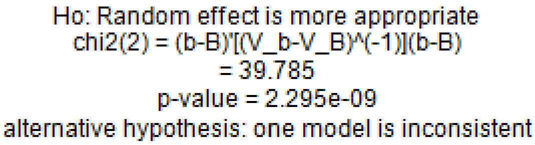
Table 3.28: Regression results of Pooled OLS model without intercept and HAC standard errors (HAC OLS estimators), Fixed effects model (Within estimators), Random effects model (Swamy and Arora (SA)-type estimators) for the LGN defaulted NHG contracts



Variables	HAC OLS	Within	SA
constant	-	0.249	0.213
standard error	-	0.034	0.019
t-value	-	7.41	11.365
p-value	-	<.0001	<.0001
Driver1	0.091	-0.180	-0.113
standard error	0.010	0.022	0.019
t-value	9.164	-8.088	-6.073
p-value	<.0001	<.0001	<.0001
Driver5	0.005	0.001	0.001
standard error	0.001	0.000	0.000
t-value	3.764	2.824	2.464
p-value	<.0002	0.005	0.014
$\hat{\sigma}^2_v$	-	-	0.001
$\hat{\sigma}^2_\mu$	-	-	0.029
R^2	0.335	0.992	-
Adjusted R^2	0.334	0.979	-
Number of observations	1255	1255	1255
Number of contracts	815	815	815

According to [Hausman, 1978] we can distinguish between fixed or random effects by performing a test based on the random effects estimator property to be inconsistent under the alternative hypothesis of correlation among the regressors $x_{i,t}$ and contract-specific effect μ_i. The Hausman type test has been performed among the within estimates and the SA Anova estimator. The result in Table 3.29 indicates the rejection of the null hypothesis.

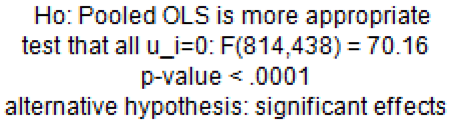
Table 3.29: Hausman test for choosing Fixed effects model or Random effects model



Ho: Random effect is more appropriate
chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 39.785
p-value = 2.295e-09
alternative hypothesis: one model is inconsistent

Thereafter, the F-test has been performed (Table 3.30) testing the assumption of fixed individual effects presence under the null hypothesis of zero unobserved individual effects. When the null hypothesis is rejected we accept that there is a significant increase in goodness-of-fit in the fixed effect model and thus the last model is more suitable than a

Table 3.30: F test for choosing Fixed effects model or Pooled OLS model

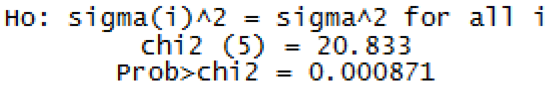


```
Ho: Pooled OLS is more appropriate
test that all u_i=0: F(814,438) = 70.16
p-value < .0001
alternative hypothesis: significant effects
```

pooled OLS. The large F statistic suggests rejection of H0 in favor of the fixed individual effects ($p < .0000$).

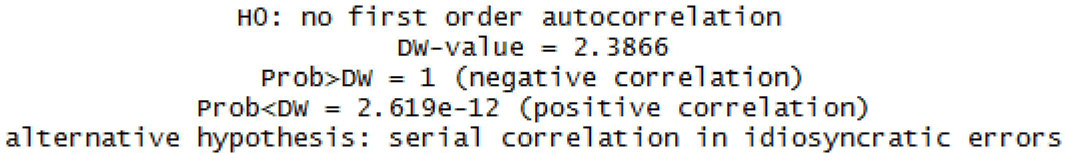
Finally, the FE model was tested regarding the heteroskedasticity and autocorrelation among the model's error terms, likewise in the pooled OLS regression. The tests are illustrated in Tables 3.31 and 3.32. The White test statistic is 20.833, value that signifies the presence of heteroskedastic residuals, while the Durbin Watson test modified for unbalanced panel data indicates that there is positive serial correlation in the regression at the 0.01 significance level. As a consequence, the Arellano robust estimators [Arellano, 1987] of the standard errors are necessary to be estimated. These fixed effects estimator's standard errors are robust to heteroskedasticity and serial correlation of arbitrary form.

Table 3.31: White test for heteroskedasticity in FE model



```
Ho: sigma(i)^2 = sigma^2 for all i
chi2 (5) = 20.833
Prob>chi2 = 0.000871
```

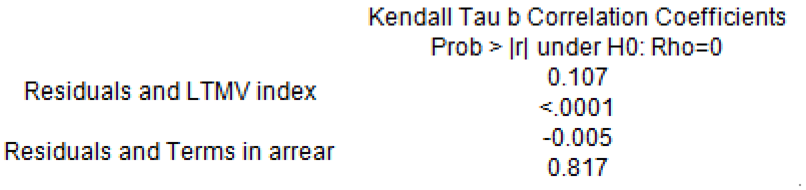
Table 3.32: Durbin-Watson test for serial correlation in FE model



```
H0: no first order autocorrelation
DW-value = 2.3866
Prob>DW = 1 (negative correlation)
Prob<DW = 2.619e-12 (positive correlation)
alternative hypothesis: serial correlation in idiosyncratic errors
```

In Table 3.33 the assumption of indepenedence among residuals and explanatory variables was investigated, testing the correlation of the model's residuals and the two explanatory variables. According to Kendal's tau correlations coefficients ([Chok, 2010]) of Driver1 and Driver5, the first has low dependece and the last one is zero-correlated with the regression's residuals.

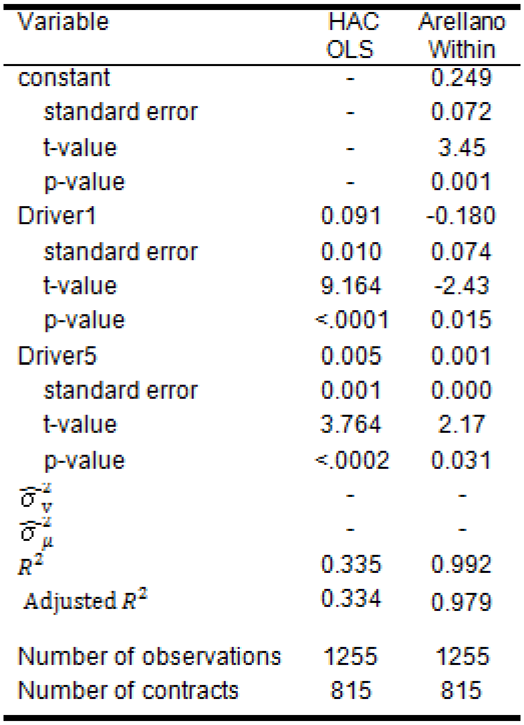
Table 3.33: The correlation analysis among regression residuals and explanatory variables for the fixed effect NHG model



	Kendall Tau b Correlation Coefficients Prob > |r| under H0: Rho=0
Residuals and LTMV index	0.107 <.0001
Residuals and Terms in arrear	-0.005 0.817

To conlude, the two models which will be compared, Pooled OLS and Fixed Effects models, for modelling the *LGN* for the NHG defaulted models are shown in Table 3.34.

Table 3.34: Regression results of Pooled OLS model without intercept and HAC standard errors (HAC OLS estimators) and Fixed effects model with Arellano standard errors (Arellano Within) for the LGN defaulted NHG contracts



Variable	HAC OLS	Arellano Within
constant	-	0.249
standard error	-	0.072
t-value	-	3.45
p-value	-	0.001
Driver1	0.091	-0.180
standard error	0.010	0.074
t-value	9.164	-2.43
p-value	<.0001	0.015
Driver5	0.005	0.001
standard error	0.001	0.000
t-value	3.764	2.17
p-value	<.0002	0.031
$\hat{\sigma}_{v}^{2}$	-	-
$\hat{\sigma}_{\mu}^{2}$	-	-
R^2	0.335	0.992
Adjusted R^2	0.334	0.979
Number of observations	1255	1255
Number of contracts	815	815

Using equations 3.2.1 and 3.2.2 the two estimated final models for the NHG contracts are given by the following Pooled model:

$$LGN_{it} = 0.091 \cdot Driver1_{it} + 0.005 \cdot Driver5_{it} + \epsilon_{it}, \tag{3.2.3}$$

and the following Fixed effect model:

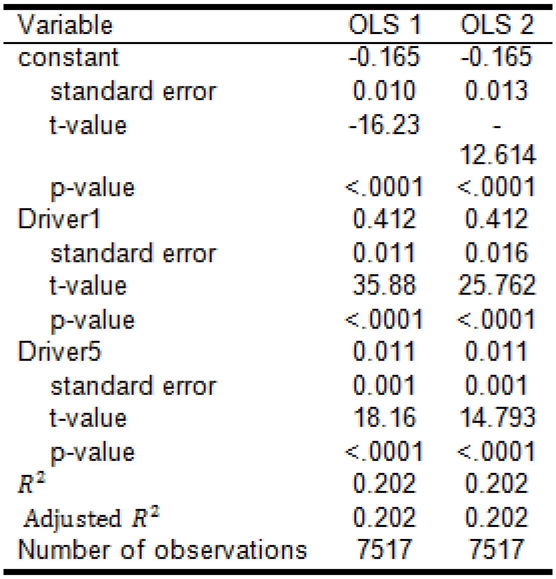
$$LGN_{it} = (0.249 + \mu_i) - 0.180 \cdot Driver1_{it} + 0.001 \cdot Driver5_{it} + v_{it}. \tag{3.2.4}$$

2.2.1.ii Non-NHG portfolio

2.2.1.ii.a Pool data modelling

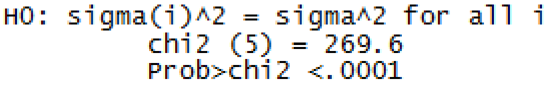
A pooled OLS regression employed with the same variables as in NHG portfolio and was tested for heteroscedasticity and autocorrelation based on White and Durbin-Watson test respectively. The detailed outputs of the regression are given in Table 3.35 and the results regarding the tests are shown in Tables 3.36 and 3.37. Similarly to the previous pooled OLS regression, the model suffers from heteroscedastic and positive serial correlated error terms at the .01 significance level. For this reason, the HAC corrected standard errors are used in the regression. The final pooled OLS model with robust standard errors is presented in Table 3.35 (OLS 2). As a final step, the Kendall's tau correlation analysis ([Chok, 2010]) among the regression's residuals and the independent variables Driver1 and Driver5 was performed. The results in Table 3.38 indicate almost no correlation for both variables, since both correlation coefficients are closed to zero.

Table 3.35: Pooled OLS regression output (OLS 1) and with HAC standard errors (OLS 2)



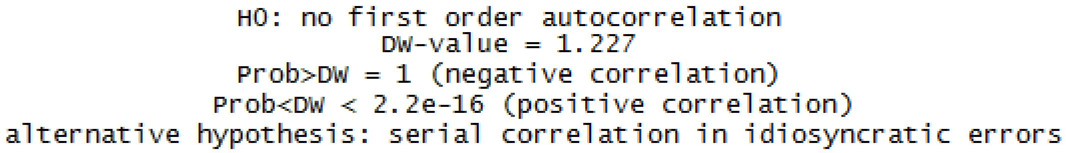
Variable	OLS 1	OLS 2
constant	-0.165	-0.165
standard error	0.010	0.013
t-value	-16.23	-12.614
p-value	<.0001	<.0001
Driver1	0.412	0.412
standard error	0.011	0.016
t-value	35.88	25.762
p-value	<.0001	<.0001
Driver5	0.011	0.011
standard error	0.001	0.001
t-value	18.16	14.793
p-value	<.0001	<.0001
R^2	0.202	0.202
Adjusted R^2	0.202	0.202
Number of observations	7517	7517

Table 3.36: White test for heteroscedasticity in pooled OLS model



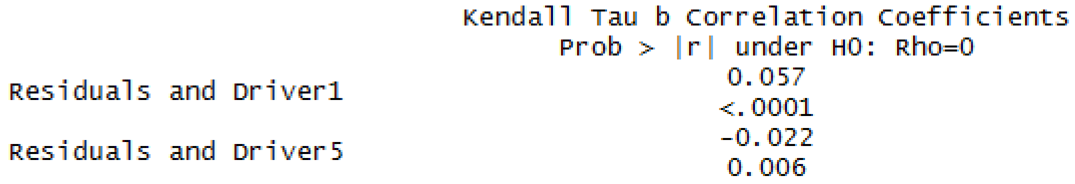
```
H0: sigma(i)^2 = sigma^2 for all i
          chi2 (5) = 269.6
          Prob>chi2 <.0001
```

Table 3.37: Durbin-Watson test for serial correlation in pooled OLS momdel



```
                H0: no first order autocorrelation
                         DW-value = 1.227
                Prob>DW = 1 (negative correlation)
              Prob<DW < 2.2e-16 (positive correlation)
alternative hypothesis: serial correlation in idiosyncratic errors
```

Table 3.38: The correlation analysis among regression residuals and explanatory variables for the OLS pooled non-NHG model

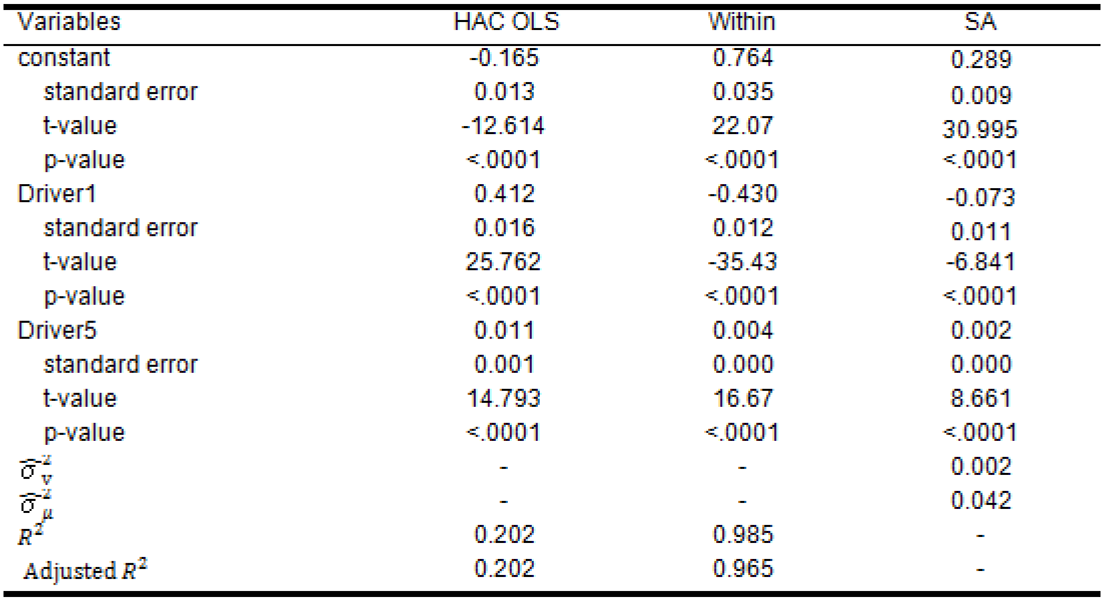


```
                                  Kendall Tau b Correlation Coefficients
                                        Prob > |r| under H0: Rho=0
                                                  0.057
Residuals and Driver1
                                                 <.0001
                                                 -0.022
Residuals and Driver5
                                                  0.006
```

2.2.1.ii.b Panel data modelling

Following the pooled model, the fixed effects and random effects panel data models were investigated. The within estimators and Swamy and Arora estimators are illustrated in Table 3.39 with the results from the pooled OLS model.

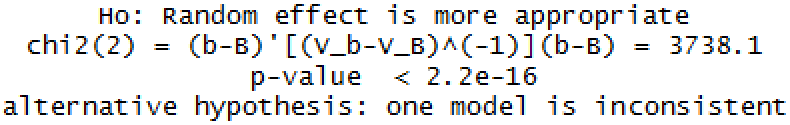
Table 3.39: Regression results of Pooled OLS model with HAC standard errors (HAC OLS estimators), Fixed effects model (Within estimators), Random effects model (Swamy and Arora (SA)-type estimators) for the LGN defaulted Non-NHG contracts



Variables	HAC OLS	Within	SA
constant	-0.165	0.764	0.289
standard error	0.013	0.035	0.009
t-value	-12.614	22.07	30.995
p-value	<.0001	<.0001	<.0001
Driver1	0.412	-0.430	-0.073
standard error	0.016	0.012	0.011
t-value	25.762	-35.43	-6.841
p-value	<.0001	<.0001	<.0001
Driver5	0.011	0.004	0.002
standard error	0.001	0.000	0.000
t-value	14.793	16.67	8.661
p-value	<.0001	<.0001	<.0001
$\tilde{\sigma}_v^2$	-	-	0.002
$\tilde{\sigma}_\mu^2$	-	-	0.042
R^2	0.202	0.985	-
Adjusted R^2	0.202	0.965	-

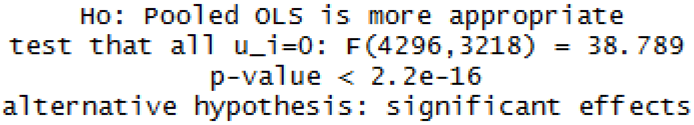
The variances of the within and Swamy-Arora's estimators were compared by means of a Hausman specification test for the choice between fixed and random effects. As can be seen in Table 3.40, the Hausman test statistic of 3738.1 resulting the rejection of the null hypothesis of uncorrelation among individual effects and regressors, and the fixed effect model preference. Consequently, the absence of fixed individuals effects in the data was examined by means of the F-test and the outcome in Table 3.41 suggests that the method which takes into consideration the ai's is the more appropriate method. In other words, the Fixed Effect model is a more representative candidate for the non-NHG portfolio LGN model.

Table 3.40: Hausman test for choosing Fixed effects model or Random effects model



```
          Ho: Random effect is more appropriate
     chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 3738.1
                   p-value  < 2.2e-16
   alternative hypothesis: one model is inconsistent
```

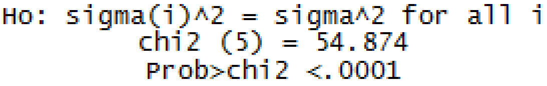
Table 3.41: F test for choosing Fixed effects model or Pooled OLS model



```
      Ho: Pooled OLS is more appropriate
  test that all u_i=0: F(4296,3218) = 38.789
               p-value < 2.2e-16
 alternative hypothesis: significant effects
```

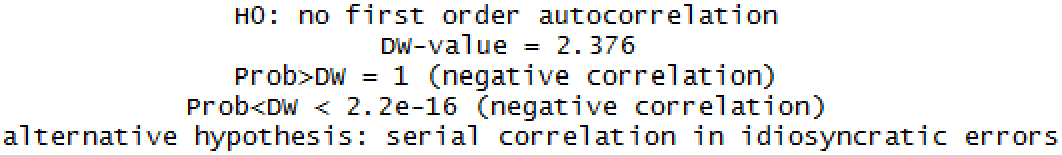
Moreover, the White test for heteroskedasticity and Durbin-Watson test for autocorrelation assumptions were performed on the FE model and the results in Tabes 3.42 and 3.43 indicate the appearance of both at the .01 significance level. As a consequence the Arellano(1987) [Arellano, 1987] robust standard errors are added to the model correcting these two issues. Thereafter, a correltion analysis among regression's residuals and explanatory variables was conducted, in order to test if the assumption of uncorrelation between them is met. The Kendall's tau correlation coefficients ([Chok, 2010]) which are presented in Table 3.44 denote very low dependence and as a result the assumption is not violated.

Table 3.42: White test for heteroscedasticity in fixed effect model



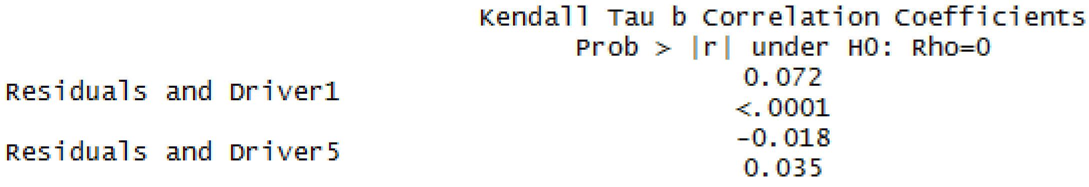
```
Ho: sigma(i)^2 = sigma^2 for all i
          chi2 (5) = 54.874
          Prob>chi2 <.0001
```

Table 3.43: Durbin-Watson test for serial correlation in fixed effect model



```
             H0: no first order autocorrelation
                       DW-value = 2.376
             Prob>DW = 1 (negative correlation)
         Prob<DW < 2.2e-16 (negative correlation)
alternative hypothesis: serial correlation in idiosyncratic errors
```

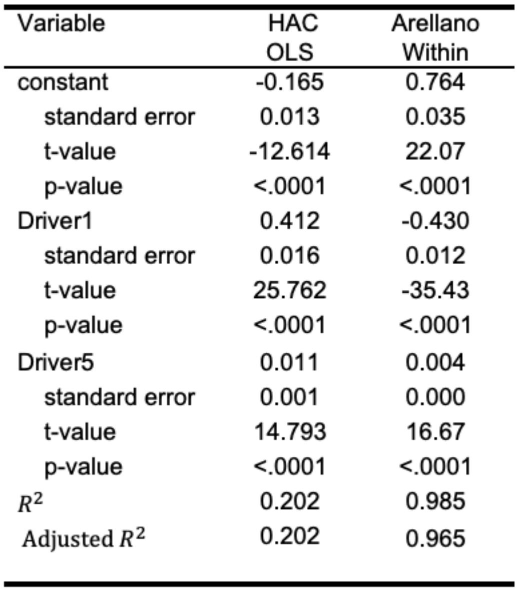
Table 3.44: The correlation analysis among regression residuals and explanatory variables for the fixed effect non-NHG model



```
                                        Kendall Tau b Correlation Coefficients
                                             Prob > |r| under H0: Rho=0
                                                      0.072
   Residuals and Driver1
                                                     <.0001
                                                     -0.018
   Residuals and Driver5
                                                      0.035
```

Overall, the pooled OLS and Fixed Effect final models for non-NHG LGN defaulted model are displayed in table 3.45 with their robust standard errors.

Table 3.45: Regression results of Pooled OLS model without intercept and HAC standard errors (HAC OLS estimators) and Fixed effects model with Arellano standard errors (Arellano Within) for the LGN defaulted Non-NHG contracts



Variable	HAC OLS	Arellano Within
constant	-0.165	0.764
standard error	0.013	0.035
t-value	-12.614	22.07
p-value	<.0001	<.0001
Driver1	0.412	-0.430
standard error	0.016	0.012
t-value	25.762	-35.43
p-value	<.0001	<.0001
Driver5	0.011	0.004
standard error	0.001	0.000
t-value	14.793	16.67
p-value	<.0001	<.0001
R^2	0.202	0.985
Adjusted R^2	0.202	0.965

In conclusion, the final models which measure the relation of $Driver1$ and $Dricer5$ with the LGN of the defaulted contracts in Non-NHG portfolio, are: Pooled OLS model:

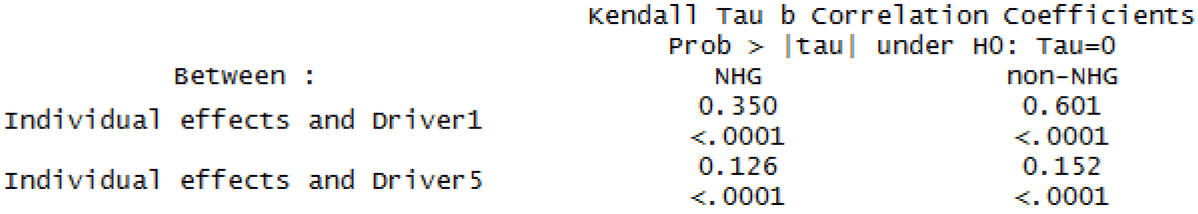
$$LGN_{(}it) = -0.165 + 0.412 \cdot Driver1_{(}it) + 0.011 \cdot Driver5_{(}it) + \epsilon_{(}it), \qquad (3.2.5)$$

and Fixed effect model:

$$LGN_{(}it) = (0.764 + \mu_i) + 0.004 \cdot Driver1_{it} - 0.430 \cdot Driver5_{it} + v_{it}. \qquad (3.2.6)$$

Thereafter the aforementioned analysis and specific after the Hausman test results, which indicate that a Fixed Effect model is more appropriate than a Random Effect model for modelling the NHG and non-NHG already defaulted contracts, the relation between the individual effects μ_i and the regressors $Driver1_{it}$ and $Driver5_{it}$ was analysed by means of a Kendall's rank correlation ([Chok, 2010]). The coefficient value 0.601 in Table 3.46 indicates the presence of high dependency among μ_i and $Driver1_{it}$ for non-NHG loans, while the tau value 0.350 shows a lower but significant correlation for μ_i and $Driver5_{it}$ for NHG loans. At the same time, the Kendall tau between μ_i and $Driver5_{it}$ for all the contracts indicates low correlation.

Table 3.46: The correlation analysis among regression individual effects and explanatory variables for the fixed effects LGN models



```
                                        Kendall Tau b Correlation Coefficients
                                            Prob > |tau| under H0: Tau=0
          Between :                          NHG                non-NHG
                                            0.350                 0.601
 Individual effects and Driver1             <.0001                <.0001
                                            0.126                 0.152
 Individual effects and Driver5             <.0001                <.0001
```

3.2.2 Model performance

In this section, in-sample and out-of-sample backtesting were implemented for both models of LGN (pooled OLS and Fixed Effects) from the previous section. Validating both models on the same dataset aim to compare their's predictive performance and decide which model gives more accurate prediction's values. The training and evaluation periods for the in-sample and out-of-sample tests are the same as in the Cure rate defaulted model performance section.

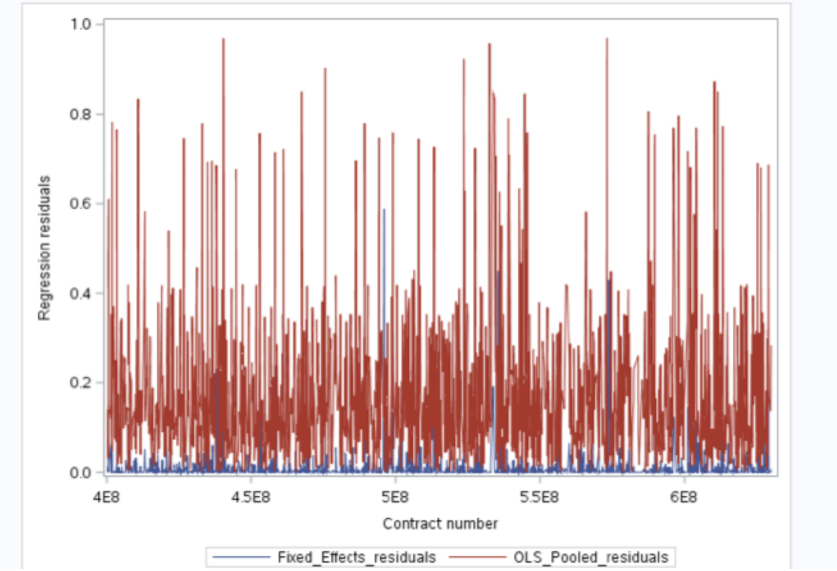
In-sample test

In this test the development dataset is the 5-year unbalanced dataset with all the non-cured contracts, and, the validation dataset consists of the last year dataset observations. Hence, the models 3.2.3-3.2.6 were used. The following tables and graphs present the residual, calibration quality and discriminatory power analysis of the two models.

Discriminatory power analysis

Figure 3.2 illustrates the absolute differences between the observed and predicted LGN of all the contracts with ultimo year 2013. As we can see the majority of LGN predicted with the Fixed effect model (blue) have lower absolute residuals than the one predicted with pooled OLS model (red). The exact percentage of the fixed effect model estimated LGN's which are closer to the realised LGN is 98%. These results are also obtained from the LGN averages over all the loans on the validation dataset. The averages of predicted LGN are displayed in table 33 and compared to the averages of observed LGN we can say that the Fixed effect model predicted more accurate LGN values than the Bank's model.

Figure 3.2: The absolute residuals from the two models on the in sample validation dataset of the LGN defaulted developed models

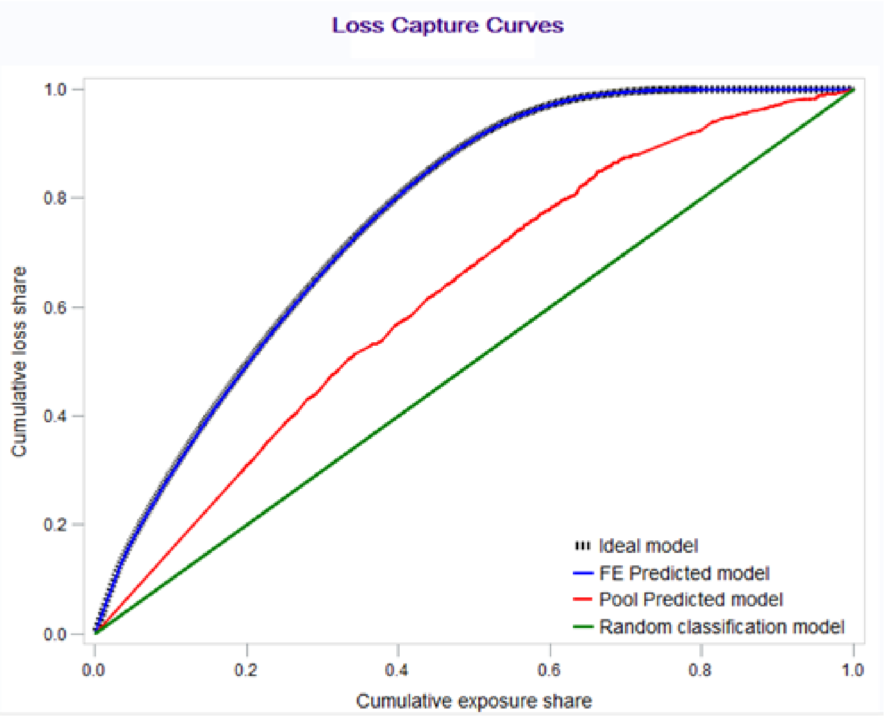


In addition, the Loss Capture Ratio (LCR) and Loss capture curve were calculated to assess the ability of the model to rank and distinguish among low and high losses [Li et al., 2009].The LCR is considered as the version of AR for continues dependent variables. Table 35 shows the results of the discriminatory power for each model based on the Loss capture ratio (LCR). The Fixed Effect model scoring 99.53% LCR and having its model curve tangent with the ideal model curve, can be considered that predicts more realistic losses than the pooled model which has the half LCR value.

Figure 3.3: The Loss Capture Ratio and Curves for each model over the in-sample validation dataset

LCR for LGN defaulted
Fixed Effect model=
99.53 %

LCR for LGN defaulted
OLS pooled model=
46.04%



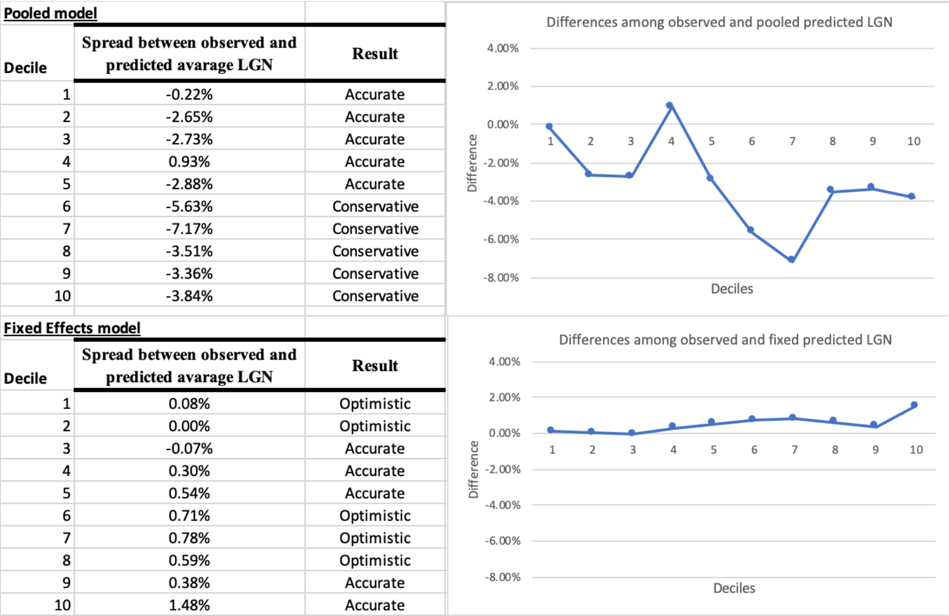
Calibration quality analysis

Moreover, the calibration accuracy of the models was tested using a t-test on the average values of the predicted LGN's for the testing dataset [Lung et al., 2003]. For each group, the t-test examines the null hypothesis that the population mean is equal to the LGN estimate. The null hypothesis is accepted when the predicted LGN lies within the confidence interval and then the model is considered "Accurate". The cases where the estimated LGN is lower than the lower confidence level (LCL) the model's performance is characterized as "Optimistic", whereas the model is called "Conservative" when the estimated LGN is higher than the upper confidence level (UCL). We performed the t-test with the same way that we performed the binomial test in the *Cure rate* model. Tables 3.47 and 3.48 show the differences between the realized and the observed losses, which were calculated for the entire validation dataset and for every decile, respectively. When the difference is positive, it means that the predicted loss is lower than the observed one, and when it is negative means the opposite. The last column of the tables indicates if the estimated LGN are "Conservative", "Optimistic" or "Accurate", based on the aforementioned definitions. The exact observed and predicted percentages of losses are confidential information and for this reason are not shown in this thesis. For a better understanding of the different predictions that the two models give, the difference

Table 3.47: Comparison between average observed and predicted LGN for the in sample validation dataset

Model	Spread between observed and predicted LGN	Result
Pooled OLS	-3.10%	Conservative
Fixed Effect logit	0.48%	Accurate

Table 3.48: The calibration accuracy for both models on the in sample validation dataset over deciles



Pooled model

Decile	Spread between observed and predicted avarage LGN	Result
1	-0.22%	Accurate
2	-2.65%	Accurate
3	-2.73%	Accurate
4	0.93%	Accurate
5	-2.88%	Accurate
6	-5.63%	Conservative
7	-7.17%	Conservative
8	-3.51%	Conservative
9	-3.36%	Conservative
10	-3.84%	Conservative

Fixed Effects model

Decile	Spread between observed and predicted avarage LGN	Result
1	0.08%	Optimistic
2	0.00%	Optimistic
3	-0.07%	Accurate
4	0.30%	Accurate
5	0.54%	Accurate
6	0.71%	Optimistic
7	0.78%	Optimistic
8	0.59%	Optimistic
9	0.38%	Accurate
10	1.48%	Accurate

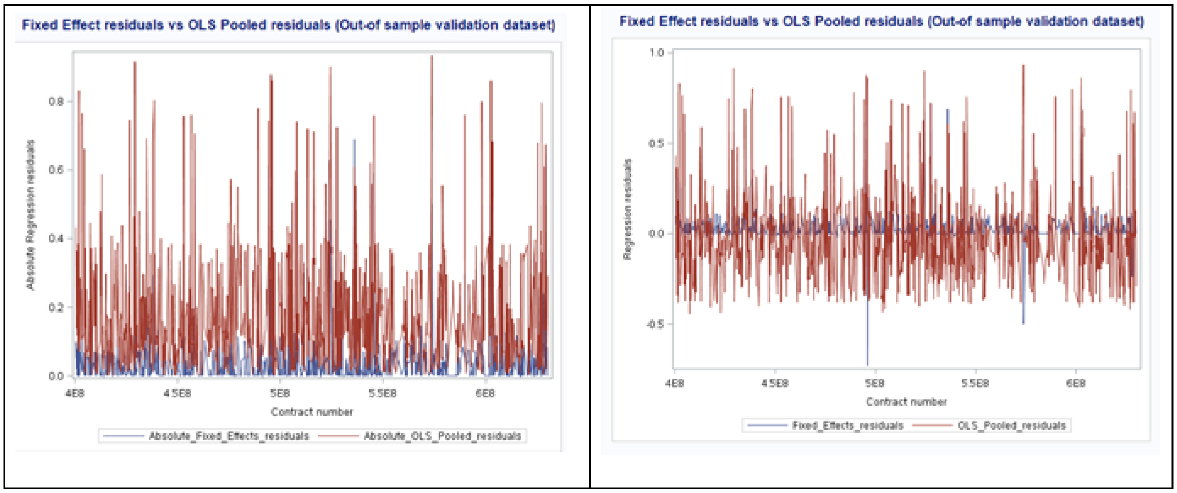
between the observed and predicted losses are illustrated graphically in Table 3.48.

From tables 3.47 and 3.48 we can see that the Pooled model estimate losses very close or higher than the real ones, in contrast with the FE model which predicts either accurate or optimistic losses. Therefore, considering that for a bank is better to be “Conservative” rather than “Optimistic”, one can result that the Pooled model is better as regards the calibration power.

Out-of-sample test

The same tests were conducted as before, but this time the development dataset consists of the loans that were in default during the first 3 years of our 5-years dataset, and from the units with remaining years were selected only the ones which were observed in the development dataset. The idea to split the dataset was based on the $Data-smplitting$ method [Giancristofaro and Salmaso, 2007], which was also used in the out-of-sample

Figure 3.4: The (absolute) residuals (left) from the two models on all the contracts of the out of sample validation dataset



test regarding the *Cure rate* model. Thus, different models were constructed based on the first three years and were applied at the contracts in the validation dataset. These tests are given by

Pooled OLS models:

$$NHG : LGN_{it} = 0.103 \cdot Driver1_{it} + 0.005 \cdot Driver5_{it} + \epsilon_{it}, \tag{3.2.7}$$

$$Non - NHG : LGN_{it} = -0.199 + 0.472 \cdot Driver1_{it} + 0.010 \cdot Driver5_{it} + \epsilon_{it}, \tag{3.2.8}$$

and the Fixed effect models:

$$NHG : LGN_{it} = (0.431 + \mu_i) - 0.110 \cdot Driver1_{it} + 0.001 \cdot Driver5_{it} + v_{it}, \tag{3.2.9}$$

$$Non - NHG : LGN_{it} = (0.842 + \mu_i) - 0.494 \cdot Driver1_{it} + 0.003 \cdot Driver5_{it} + v_{it}. \tag{3.2.10}$$

After applying the aforementioned models to the testing dataset, the performance of the two models was examined and compared.

Discriminatory power analysis

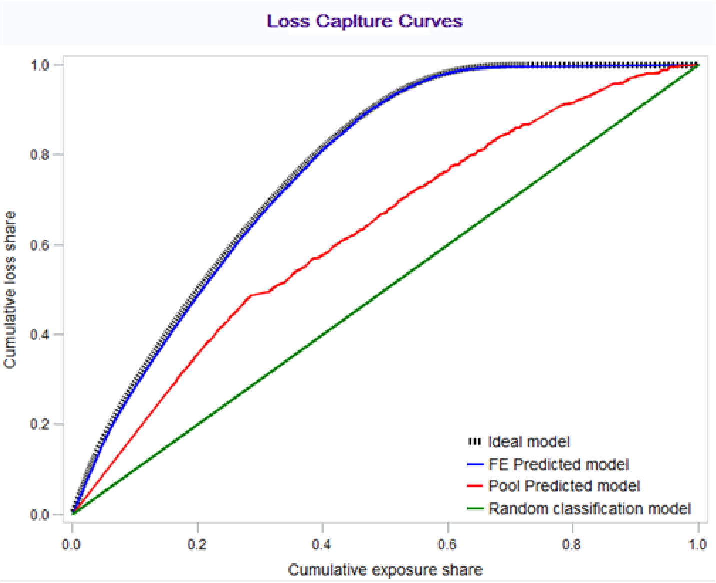
The two figures in the table 3.4 demonstrate the differences between the observed and predicted LGN of the contracts in the validation dataset. Similarly to the in-sample validation, these figures indicate that the Fixed effect model'w residuals (blue) are lower than the pooled OLS residuals (red). In more details, about the 86% of the contracts in the validation dataset got a more accurate estimation of LGN when the individual fixed effects added to the model.

The discriminatory power of the models was also tested constructing the curves with cumulative share of realized EAD on the horizontal axis and cumulative share of the

Figure 3.5: The Loss Capture Ratio and Curves for each model over the out-of-sample validation dataset

LCR for LGN defaulted
Fixed Effect model=
97.83 %

LCR for LGN defaulted
OLS pooled model=
47.76%

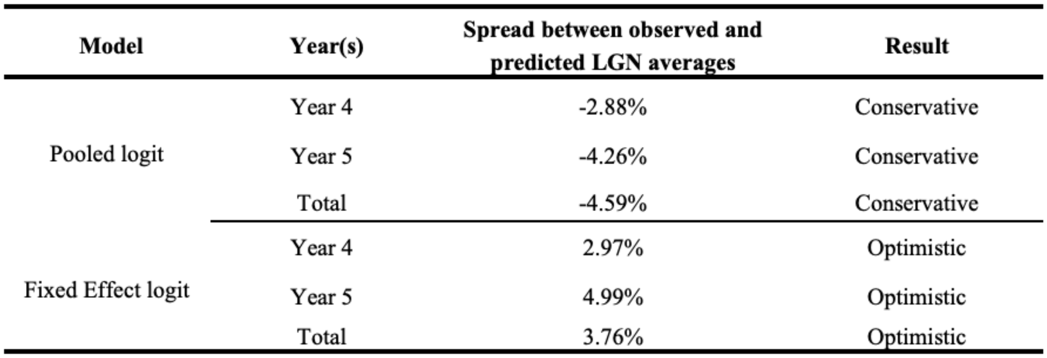


realized loss on the vertical axis [Li et al., 2009]. The results in Table 39 are surprisingly the same as in the in-sample test, indicating that the discriminatory ability of the two models is the same regarding the data that are used to build it and data which were not included in the development dataset. The FE model manages to distinguish among almost all the different levels of losses better than the pooled model. This is evidenced by the LCR outputs for the two models, where for the FE model is almost 100% and for the pooled model approximately 50%.

Calibration quality analysis

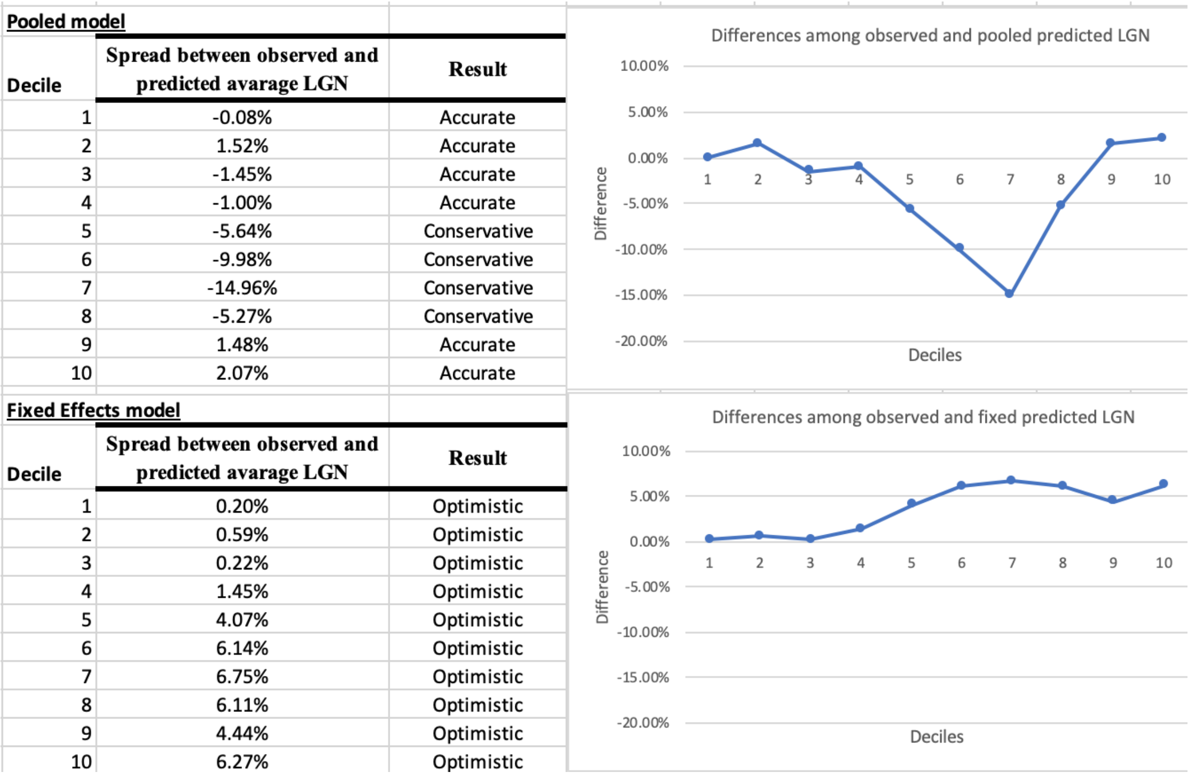
Furthermore, the model's calibration accuracy was examined through T-test over the averages per validation year and deciles [Lung et al., 2003]. From table 3.49 we see that the pooled model estimates loss for every year higher than the realized loss and at the same time the fixed effect model behaves opposite giving optimistic predictions for the LGN. When we analyse in more detail over deciles (Table 3.50), the FE model kept its "optimistic" character across all the deciles, whereas the Pooled model predicted either accurate or conservative losses. As we mentioned on the backtesting section of the Cure rate model, a bank prefers a model with conservative estimations rather than optimistic. Based on this, we may conclude that the Pooled model would be considered better for a bank than the FE model.

Table 3.49: The calibration accuracy for both models on the out-of-sample validation dataset over years



Model	Year(s)	Spread between observed and predicted LGN averages	Result
Pooled logit	Year 4	-2.88%	Conservative
	Year 5	-4.26%	Conservative
	Total	-4.59%	Conservative
Fixed Effect logit	Year 4	2.97%	Optimistic
	Year 5	4.99%	Optimistic
	Total	3.76%	Optimistic

Table 3.50: The calibration accuracy for both models on the out of sample validation dataset over deciles



Pooled model

Decile	Spread between observed and predicted avarage LGN	Result
1	-0.08%	Accurate
2	1.52%	Accurate
3	-1.45%	Accurate
4	-1.00%	Accurate
5	-5.64%	Conservative
6	-9.98%	Conservative
7	-14.96%	Conservative
8	-5.27%	Conservative
9	1.48%	Accurate
10	2.07%	Accurate

Fixed Effects model

Decile	Spread between observed and predicted avarage LGN	Result
1	0.20%	Optimistic
2	0.59%	Optimistic
3	0.22%	Optimistic
4	1.45%	Optimistic
5	4.07%	Optimistic
6	6.14%	Optimistic
7	6.75%	Optimistic
8	6.11%	Optimistic
9	4.44%	Optimistic
10	6.27%	Optimistic

3.3 The LGD model

Finally, we developed two approaches regarding a LGD model with pooled and panel data. As we mentioned at the beginning of this chapter the LGD model has been defined as

$$LGD = (1 - Cure) \cdot LGN$$

Having derived both underlying components in Section 1 and 2, for the defaulted loans, we evaluate in this section, the in-sample and out-of-sample predictive performance given both approaches.

3.3.1 Model performance

In-sample test

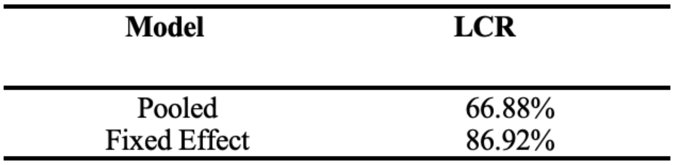
The development and validation datasets are constructed in the same way as before. Hence, the original 5-years unbalanced dataset with the cure and non-cure defaulted contracts was used, to build the *Cure rate* and *LGN* models. The models are the one that we found previously and are given from equations 3.1.4- 3.1.5 and 3.2.3-3.2.6. Then, the resulted models were applied to the defaulted loans that were observed during the year of the dataset.

Discriminatory power and Calibration quality analysis

The classification ability of the LGD models among the different level of losses was tested by means of the Loss Capture Ratio (LCR) [Li et al., 2009]. The results in table 3.51 indicate that the ability of the FE model to distinguish among low and high losses is better than the ability of the pool model, since they score almost 87% and 67% LCR, respectively. Moreover, regarding the calibration power, the validation dataset was grouped into deciles and therefore the two models were tested not only on the entire validation dataset but also on each decile. A t-test was performed over the averages and we categorised again the model as "Conservative", "Optimistic" or "Accurate", as before [Lung et al., 2003].

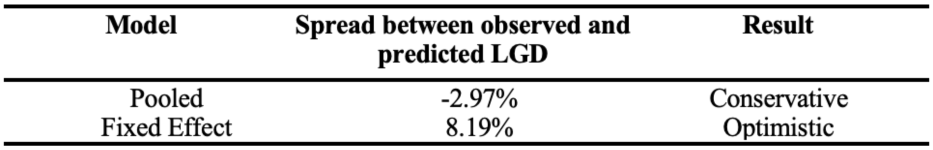
The two models compute different predictions as the FE model in most of the deciles in Table 3.53 gives optimistic estimations, in contradiction with the conservative and accurate predictions of the pool model. This also can be seen from Table 3.52 where the accuracy of the two models were calculated over the entire validation dataset. From the graphs in Table 3.52, we can see that both models predict losses close to the realised ones. But, when it comes to the 9th and 10th deciles the FE model gives estimations that are very different from the real values of losses.

Table 3.51: The Loss Capture Ratio for each model over the entire in-sample validation dataset



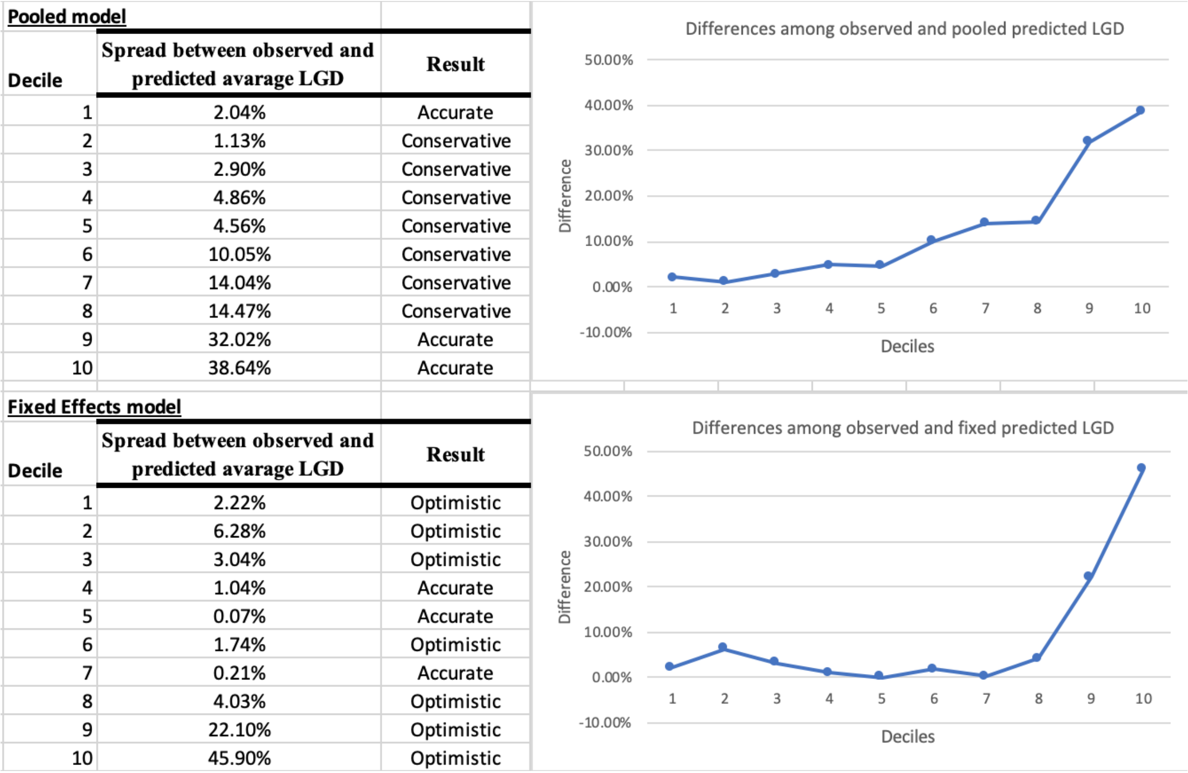
Model	LCR
Pooled	66.88%
Fixed Effect	86.92%

Table 3.52: T-test results for the entire in-sample validation dataset



Model	Spread between observed and predicted LGD	Result
Pooled	-2.97%	Conservative
Fixed Effect	8.19%	Optimistic

Table 3.53: T-test results for the in-sample validation dataset per decile



Pooled model

Decile	Spread between observed and predicted avarage LGD	Result
1	2.04%	Accurate
2	1.13%	Conservative
3	2.90%	Conservative
4	4.86%	Conservative
5	4.56%	Conservative
6	10.05%	Conservative
7	14.04%	Conservative
8	14.47%	Conservative
9	32.02%	Accurate
10	38.64%	Accurate

Differences among observed and pooled predicted LGD
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
-10.00%
Difference
1
2
3
4
5
6
7
8
9
10
Deciles

Fixed Effects model

Decile	Spread between observed and predicted avarage LGD	Result
1	2.22%	Optimistic
2	6.28%	Optimistic
3	3.04%	Optimistic
4	1.04%	Accurate
5	0.07%	Accurate
6	1.74%	Optimistic
7	0.21%	Accurate
8	4.03%	Optimistic
9	22.10%	Optimistic
10	45.90%	Optimistic

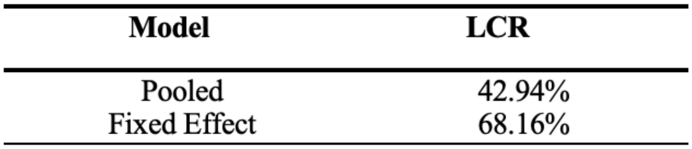
Out-of-sample test

For this test, the first 75% of the entire dataset was used to build the cure rate and LGN models. This ratio corresponds to the defaulted contracts across the first 3 years of out 5-years dataset and it is based again in the *Data − splitting* method ([Giancristofaro and Salmaso, 2007]). The resulting models are the models that we found in the previous Out-of-sample tests and are given in equations 3.1.11-3.1.12 and 3.2.7-3.2.10. Both models were applied to the remaining loans from the original dataset. Combining the results from the two models the LGD model were obtained.

After assessing the discriminatory power of the panel LGD and pool LGD models we concluded to the same inference as in the in-sample test. First, the Loss Capture Ratio (LCR) was computed for the two models [Li et al., 2009]. The FE model has better discriminatory performance since scores a higher LCR (Table 3.54). When it comes to the ability of the models to compute predictions as much closer to the real ones, both models fail two compute accurate predictions according to the results from the t-test [Lung et al., 2003] in tables 3.55 and 3.56. The model which considers the fixed contract-specific effects gives mostly optimistic results, when the t-test was applied on the entire validation dataset and in each decile. On the other hand, the model that ignores the cross-sectional effects estimates accurate or conservative predictions, with estimated amount of losses same or higher than the realized ones.

Discriminatory power analysis

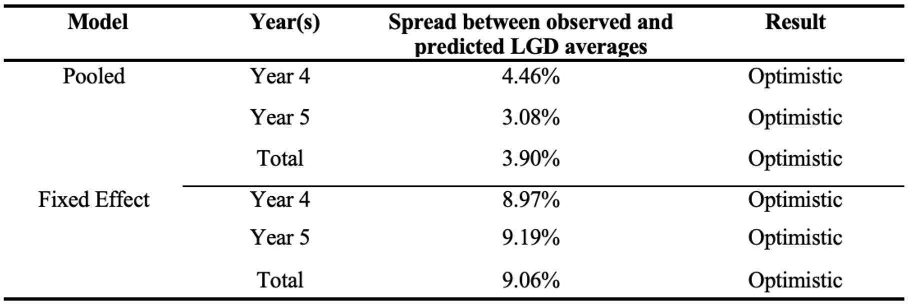
Table 3.54: The Loss Capture Ratio for each model over the out-of-sample validation dataset



Model	LCR
Pooled	42.94%
Fixed Effect	68.16%

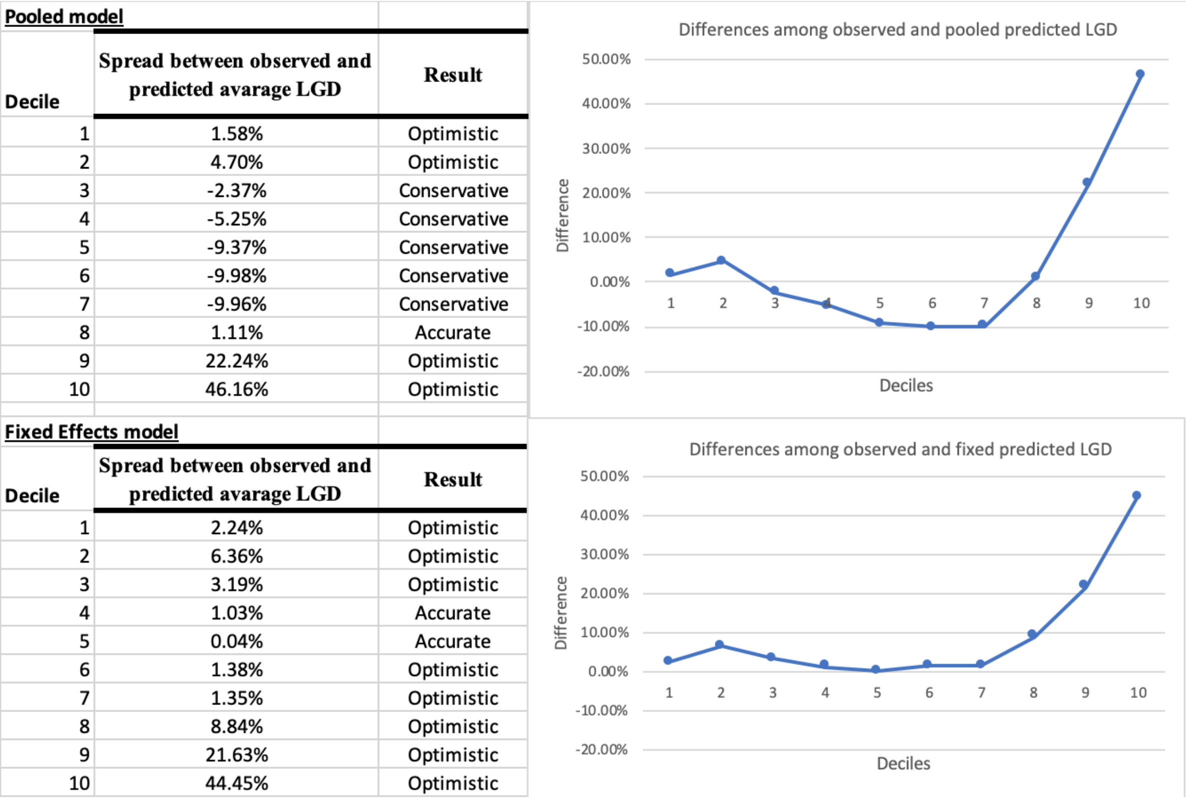
Calibration quality analysis

Table 3.55: T-test results for the entire out-of-sample validation dataset



Model	Year(s)	Spread between observed and predicted LGD averages	Result
Pooled	Year 4	4.46%	Optimistic
	Year 5	3.08%	Optimistic
	Total	3.90%	Optimistic
Fixed Effect	Year 4	8.97%	Optimistic
	Year 5	9.19%	Optimistic
	Total	9.06%	Optimistic

Table 3.56: T-test results for the out-of-sample validation dataset per decile



Pooled model

Decile	Spread between observed and predicted avarage LGD	Result
1	1.58%	Optimistic
2	4.70%	Optimistic
3	-2.37%	Conservative
4	-5.25%	Conservative
5	-9.37%	Conservative
6	-9.98%	Conservative
7	-9.96%	Conservative
8	1.11%	Accurate
9	22.24%	Optimistic
10	46.16%	Optimistic

Differences among observed and pooled predicted LGD
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
-10.00%
-20.00%
Difference
1
2
3
4
5
6
7
8
9
10
Deciles

Fixed Effects model

Decile	Spread between observed and predicted avarage LGD	Result
1	2.24%	Optimistic
2	6.36%	Optimistic
3	3.19%	Optimistic
4	1.03%	Accurate
5	0.04%	Accurate
6	1.38%	Optimistic
7	1.35%	Optimistic
8	8.84%	Optimistic
9	21.63%	Optimistic
10	44.45%	Optimistic

The conclusion regarding the model assessment can be found in the next Chapter.

4

Conclusion

4.1 Conclusion

In this thesis we have constructed and compared different models which estimate a Loss given default (LGD) for a mortgage portfolio. Our dataset was unbalanced while every observation represents a contract-year information. This cross-sectional time-series dataset form allows us to build either a pooled or a panel data model. For the panel data, it was necessary to develop two different models: one with fixed unobserved cross-section specific effects and another with random unobserved effects. A fixed-effects model tests the individual differences in intercepts without making any assumptions about the correlation among the individual effects and the regressors. However, the random effects model assumes that individual differences are part of the error term and independent of the regressors. Then, by using some statistical tests, one can choose the most appropriate for the data.

The LGD model relied on two underlying models: a Cure rate and a Loss given No-cure (LGN) model. The Cure rate model is a discrete choice model, and it calculates the proportion of defaulted counterparties curing from default. The nature of the other model is linear, and it estimates the amount of money lost from non-cured defaulted loans with the continuous variable LGN as its dependent variable.

For both aforementioned models, linear and non-linear, bivariate analysis was employed, in order to assess the association among the dependent and the possible explanatory variables. Considering the results of this analysis support by expert opinion, we concluded that the so-called Driver5 is a significantly variable and forms the best explanatory regressor in classifying cure defaulted loans. As regards to the LGN model, we found that the Driver1 together with the Driver5 are the most appropriate variables to describe the loss from non-cured contracts. After deciding the covariates, the pooled and panel data models were constructed for the Cure rate, LGN, and ultimately for the LGD.

First, we focused on the estimation of the Cure rate model, where we used the traditional method of logistic regression for the pooled data assumption. As expecetd, the resulting parameter estimator was negative underlying the negative relation of Driver5 with the probability for the client to cure. Regarding the panel data with fixed contract-specific effects, the Firth's bias-reduce logistic regression technique was applied, as Kunz et al. ([Kunz et al., 2017]) proposed. In this approach, the MLE were obtained after the first-order bias was removed from the score functions of the unknown parameters. The estimators derived from an iteratively re-weighted least squares (IWLS) algorithm [Kosmidis, 2007]. The Random effects probit model was constructed for the panel data with random unobserved individual effects, and the MLE's were derived by using the conditional joint density of the error terms upon the random effects. Afterwards, for deciding the most appropriate model to represent our data, two different statistical tests were employed: A Wald test, based on the assumption that random individual effects are independent of explanatory variables, and a Hausman test [Hausman, 1978], which uses the differences between the fixed effect biased reduction logit MLE and the usual logit MLE. From these tests we found that the fixed effect model appears to be more suitable for the data than the random effects or the pooled logit model.

Next, the LGN model was constructed using the Driver1 and Driver5 as covariates. Assuming that the data are pooled cross-sectional time-series, an OLS linear regression analysis was employed. From the obtained regressors estimators we found that a high Driver1 ratio results to a higher loss as well as a high Driver5 is related to a higher loss. After the pooled OLS regression analysis, the relation of the explanatory variable to the independent variables has been examined through a panel data analysis with fixed and random effects. For the fixed effects linear model, we used the within effect estimation method to derive the estimators, whereas for the random effects model the Swamy and Arora (SA)-type estimators were computed, which is a feasible generalized least square (FGLS) method based on the within and between estimation approaches. The variances of the within and Swamy-Arora's estimators were compared by means of a Hausman specification test for the choice between fixed and random effects. Thereafter, the absence of fixed individuals effects in the data was examined by means of an F-test. The results of these tests have suggested that the method which does not ignore the contract-specific effects and assumes that these unobderved effects are fixed for every contract is the most appropriate model for the LGN.

Overall, the above analyses have indicated that, for both Cure rate and LGN models, the assumption that individual fixed effects exist among the data is more apropriate than assuming that these effects are random. This result was expected, since we aimed to construct a model based on the behaviour and characteristics of the N specific contracts that we have in our dataset and compute the different unobserved effect for each one of them. We were not interested into building a model based on a randomly selected set of N individuals from a large population.

Following these steps, the performance of the pooled and panel data fixed effect (FE) models was compared through in-sample and out-of-sample backtesting tests, in order to derive their predictive power on the contracts used to build them and on a different dataset than the development one. All loans that were observed in the dataset to be in default during a given year were used for the model validation process (validation dataset). In contrast to the in-sample method, the data that the out-of-sample test uses to validate the performance of the model are not included in the development dataset. Using the Data-splitting method, the entire dataset is split into two sub-datasets, obtaining the development and testing samples. Therefore, models were constructed, based on the defaulted contracts across a specific time window and the performance of these models was tested on the contracts which were out of this time window. The validation datasets of the tests were used to assess the discriminatory power and calibration ability of the models.

We found that the two Cure rate models, pooled and FE, are not sample-specific given the almost identical results from the in-sample and out-of-sample performance tests for both models. Moreover, the FE model was found to be better than the pooled model in distinguishing cured from non-cured loans. This is evidenced from the higher overall accuracy, number of correct predicted cured contracts and number of correct estimated non-cured loans of the first model over the pooled one when the threshold probability c belongs to . Additionally, the high ability of the model which considers the loan-specific fixed effects in order to differentiate the defaulted contracts among cured and non-cured, is deduced from the very high output values of the Kolmogorov-Smirnov statistics, Accuracy Ratios and Area under the ROC curves. Specifically, all three statistics score very close to 1.00 for the FE model, while the pooled model scores are much lower than 1.00. With regard to the ability of the models to predict cure rates really close to the observed ones the two models perform very differently. The FE model based on the Binomial test and Hosmer-Lemeshow test results has an overall “optimistic” character predicting most of the times higher cure probabilities than the realized. On the contrary, the pooled logit model estimates either accurate probabilities or lower cure ratios than the real one (conservative estimations). For a company, it is better to predict that less clients will cure and calculate higher loss than the realized, rather than expecting a higher cure rate than the true one and be in the unpleasant position to find out that its loss will be higher than expected. Consequently, under the condition that the conservative estimations are not “bad” estimations, the Pooled model performed better than the FE model with respect to their calibration ability.

Additionally, the performance of the Pooled and FE LGN models was evaluated and, as before, the results from the in-sample and out-of-sample tests were not significantly different. The ability of the models to rank and distinguish among low and high losses was assessed by means of the Loss Capture Ratio (LCR) and Loss capture curve. The Fixed Effect model scoring almost 100% LCR, can be considered to predict more realistic losses than the pooled model which has about the half LCR value. Furthermore, the calibration accuracy of the models was examined through the use of a T-test over the averages of the observed and estimated losses per validation year and deciles of the validation data yielding inferences similar with the Binomial test results of the Cure rate models. The pooled model has estimated either accurate or conservative losses, however, by adding the contract-specific effects in the model and derive the fixed effect model the predictions for the LGN becoming optimistic most of the times. As we mentioned earlier, a bank prefers a model with conservative estimations rather than optimistic ones. Based on this, we concluded that, for the purposes of a bank, the Pooled model is considered better than the FE model as far as the calibration power is concerned.

Finally, by aggregating the Cure rate and the LGN models, we constructed the LGD models for pooled and panel data. The classification ability of the model among the different level of losses that were observed from the defaulted loans among the 5-year period was improved when the fixed contract effects were taken into consideration instead of being ignored. In more details, the LCR for the FE model was found to be about 87% and 68%, whereas for the Pooled model was approximately 67% and 43%, for the in-sample and out-of-sample tests, respectively. When it comes to the models' ability to estimate the most accurate predictions, the two models, once again, behaved differently. The Pooled model estimates the same or higher amount of losses than the realized ones whilst the model with the fixed contract-specific effects gives a prediction for losses which is the same or lower from the observed one. Therefore, as expected from the performance of the Cure rate and LGN models the Pooled LGD model is preferable than the FE model given its accurate and conservative behavior.

To summarize, we can conclude that by adding the unobserved contract-specific effects in all three models (Cure rate, LGN and LGD) the form of the resulting model and its performance has changed, particularly when compared to model that ignores these effects. First, the discriminatory power of the FE models is better than the Pooled model classification ability. Second, a surprising result is that, in all cases, the two models performed differently as far as their calibration ability is concerned. On the one hand, the model which is based on the traditional linear and logistic techniques, the Pooled model, is a more conservative approach and on the other hand the panel data Fixed effect model gives more optimistic predictions. Consequently, when a bank prefers a more conservative attitude, the Pooled model has preferable estimations.

4.2 Shortcomings of the proposed approach

The panel data analysis does not only have advantages, but also includes certain shortcomings. An important drawback is the fact that it cannot be used for forecasting predictions when the cross-sectional units are not included in the dataset that was used to build the model. If a new individual appears in the dataset that is used to test the performance of the model or for forecasting reasons the data analysts, this individual will not be able to apply the panel data model, since the individual random or fixed effects will be missing resulting to the pooled model.

Moreover, the construction of a panel data model requires more complex calculations than the pooled data model, especially when we refer to the logistic models. Consequently, more time is required for running the algorithms and deriving the parameters estimations for the first model. In addition, it is of vital importance for the researcher to be able to understand the tests and interpret their results when it comes to the decision of whether fixed or random effects model represents better the data.

4.3 Future work

In this thesis, we have constructed some models and then we assessed their performances using the data-splitting validation method. According to [Giancristofaro and Salmaso, 2007], this method splits randomly the original dataset into two sub-datasets with the splitting portion to lie between 2/3 and 3/4. The biggest sub-dataset corresponds to the data that will be used to develop the model and this model will be applied to the other dataset in order to validate it. [Giancristofaro and Salmaso, 2007] have also mentioned the repeated data-splitting technique. This method repeats the steps from the data-splitting method several times, getting different sub-datasets in each iteration. Thus, this method is more accurate than the previous one (Harrell et al., 1996) [Harrell et al., 1996]. Therefore, if time allows, it will be interesting to examine the model's performance by applying the repeated data-splitting validation technique and compare it with the accuracy of the data-splitting method.

When choosing to apply or follow the standard statistical practice, it is necessary to select a single model that fits the data reasonably well among a class of models and make predictions and inferences as if the selected model is the best for the data. But is this the best approach? If another model also fits well the data but leads to different statistical inferences, like effect sizes or predictions, then, it is risky to derive conclusions based only on the results of the first or second model. Therefore, the standard statistical techniques ignore the model's uncertainty and give over-confident results, which might not lead to the maximum predictive coverage [Hoeting et al., 1999].

Fortunately, Bayesian model averaging (BMA) deals with the problem of uncertainty in model selection. This approach suggests considering different models instead of using only one. It assumes that the posterior distribution of the quantity of interest given the data is the average of its posterior distributions under each of the different considered models, multiplied with the posterior probability of every model. The quantity of interest might be a parameter or a future observation. According to [Hoeting et al., 1999] and [Madigan and Raftery, 1994], BMA improves the predictive performance when it is compared with the predictive ability of a single model. One of the obstacles that might rise from this method is that the number of the considered models that fit well the data is enormous, and this may lead to difficult computations. Two of the most popular approaches that provide a panacea to this problem by eliminating the number of considered models are the Occam's window method and the Markov chain Monte Carlo model composition (MC_3) [Hoeting et al., 1999]. The first approach takes the average of a subset of the models, and the second takes the average of a function based on a Markov chain over every well-fitted model.

Appendix

A. Assumption Tests

A1. Cure rate model

LM test for heteroscedasticity on a binary response model

Heteroscedasticity is the existence of different variances between the regression disturbances across time and individuals. Assuming homoscedasticity (the same variance for all the error terms) while there are heteroscedastic error terms will lead to consistent but inefficient estimates. Therefore, it is vital to test for homoscedasticity instead of blindly assume it (Blatagi, 2005).

If the regression errors terms suffer from heteroskedasticity then their variance depends on exogenous variables z_{it} like that

$$\mathbb{V}(\epsilon_{it}) = \pi^2/3h(z'_{it}\alpha) \tag{4.3.1}$$

where $\pi^2/3$ is the variance of a logit model and h is some function $h > 0$ with $h(0) = 1$ and $h'(0) \neq 0$ (Verbeek, 2004, p.358) . Note that z_{it} should not include a constant and its dimension is J.

The Lagrange multiplier test statistic examine the null hypothesis that $alpha = 0$ and thereofre the residuals in the binary response model are homoscedastic is given by $LM = NR^2$, where the measure R^2 will resulted from the regression of ones upon the variables $\hat{\epsilon}^G_{it}x'_{it}$ and $(\hat{\epsilon}^G_{it} \cdot x'_{it}\hat{\beta})z'_{it}$ with the term $\hat{\epsilon}^G_{it}$ denoting the generalized residual of the regression model. Under the null hypothesis the test statistic LM is Chi-squared distributed with P degrees of freedom.

A2. LGN model

White test for heteroscedasticity on a linear model

The White test is using the results of an auxiliary regression analysis: the regression of squared residuals on a constant , all first moments, second moments and cross products of the original regressors The number of the auxiliary regressors without the intercept is P.

This test does not make any assumption about the structure of the heteroscedasticity. The null hypothesis of homoscedasticity is $Ho : {sigma_i}^2 = sigma^2$ for all i against the alternative hypothesis H1 of heteroscedasticity. The test statistic is the product of the R^2 value and sample size: LM = nR^2 and is assimptotically distributed as Chi-squared with P degrees of freedom.

Durbin-Watson test for serial correlation on a linear model

Autocorrelation or serial correlation among the regression error terms holds when the residuals are not independent from each other but are correlated instead [Verbeek, 2004, p97]. One of the forms of autocorrelation is the so called first order autocorrelation where the regression error term is expressed as

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + v_{it} \tag{4.3.2}$$

with $\mid \rho < 1$, ϵ_{it} be the regression error terms for every cross-sectional unit i on period t and v_{it} denoting a no autocorrelated error term with zero mean and constant variance [Verbeek, 2004, pp98–101, 357]. When ρ is different from zero the residual from individual i over time t is linearly dependent with the its residual from the previous time period $t - 1$ and therefore autocorrelation exists.

According to [Verbeek, 2004, pages 102,357] one of the most popular tests for testing the sign of ρ is the Durbin-Watson test which was introduced from Bhargava et al (1982)[Bhargava et al., 1982] with null hypothesis $H0 : \rho = 0$. The test statistic is as follows

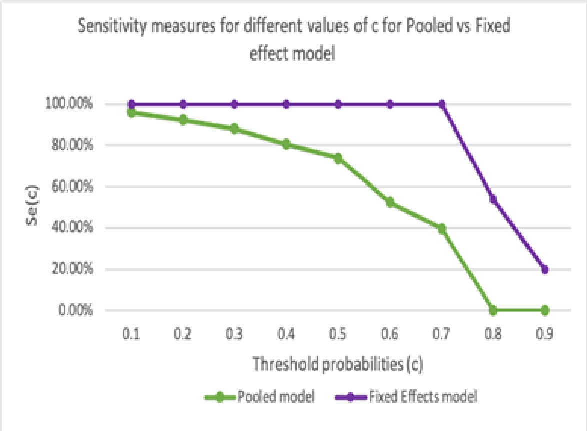
$$dw = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T_i} \hat{\epsilon}_{it} - \hat{\epsilon}_{i,t-1}{}^2}{\sum_{i=1}^{N}\sum_{t=1}^{T_i} \hat{\epsilon}_{it}^2} \tag{4.3.3}$$

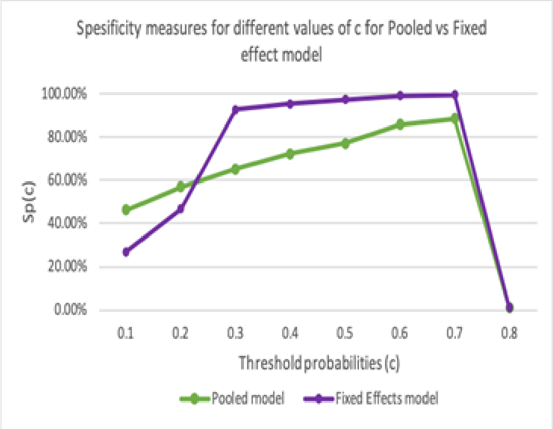
,

where ϵ_{it} represents the OLS residual for the pooled model and the regression residual from the within transformed model regarding the fixed effect model. The distribution of dw depends upon the size of N, Ti and the size and values of regressors and as a result it is difficult to obtain general critical values for the test statistic dw. Fortunately, Bhargava et al (1982)[Bhargava et al., 1982] showed that the relation of dw with the estimated ρ , $\hat{\rho}$, is given by $dw \approx 2 - 2\hat{\rho}$. Thus, when the dataset has a very large number of cross-sectional units N, the model does not suffer from serial correlation and ρ is close to 0, if the value of dw is close to 2. However, when dw is much smaller or much larger than 2, it indicates that the error terms are positive correlated to each other ($\rho > 0$) or negative ($\rho < 0$).

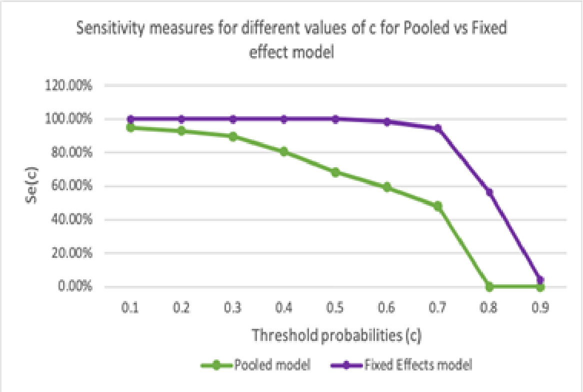
B. Plots of sensitivity and specificity measures across different threshold probabilities c, for pooled logit and fixed effect logit models

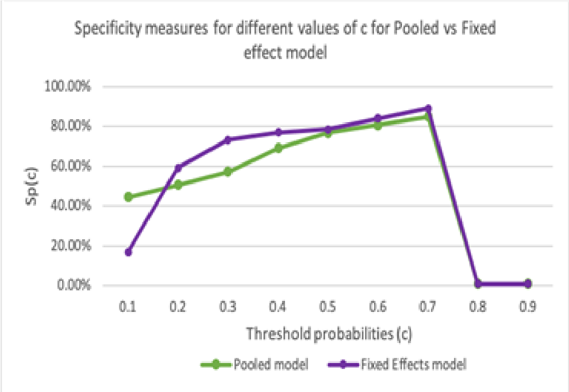
In-sample test





Out-of-sample test





References

[Abrevaya, 1997] Abrevaya, J. (1997). The equivalence of two estimators of the fixed-effects logit model. *Economics Letters*, 55(1):41–43.

[Amemiya, 1971] Amemiya, T. (1971). The estimation of the variances in a variance-components model. *International Economic Review*, pages 1–13.

[Arellano, 1987] Arellano, M. (1987). Practitioners'corner: Computing robust standard errors for within-groups estimators. *Oxford bulletin of Economics and Statistics*, 49(4):431–434.

[Baltagi, 2005] Baltagi, B. (2005). *Econometric analysis of panel data, 3rd Edition.* John Wiley & Sons.

[Baltagi, 2013] Baltagi, B. (2013). *Econometric Analysis of Panel Data, 4th Ed.* Wiley.

[BCBS, 2006] BCBS (2006). *International Convergence of Capital Measurement and Capital Standards: a Revised Framework.*

[Bester and Hansen, 2009] Bester, C. A. and Hansen, C. (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics*, 27(2):131–148.

[Bhargava et al., 1982] Bhargava, A., Franzini, L., and Narendranathan, W. (1982). Serial correlation and the fixed effects model. *The Review of Economic Studies*, 49(4):533–549.

[Björklund, 1985] Björklund, A. (1985). Unemployment and mental health: some evidence from panel data. *Journal of Human Resources*, pages 469–483.

[Breusch and Pagan, 1980] Breusch, T. S. and Pagan, A. R. (1980). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pages 1287–1294.

[Butler and Moffitt, 1982] Butler, J. S. and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica: Journal of the Econometric Society*, pages 761–764.

[Cecchetti, 1986] Cecchetti, S. G. (1986). The frequency of price adjustment: a study of the newsstand prices of magazines. *Journal of Econometrics*, 31(3):255–274.

[Chamberlain, 1980] Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225–238.

[Chamberlain, 1984] Chamberlain, G. (1984). Panel data. *Handbook of econometrics*, 2:1247–1318.

[Chok, 2010] Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data.* PhD thesis, University of Pittsburgh.

[Dhaene and Jochmans, 2015] Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030.

[Diamond, 1992] Diamond, G. A. (1992). What price perfection? calibration and discrimination of clinical prediction models. *Journal of clinical epidemiology*, 45(1):85–89.

[Engelmann et al., 2003] Engelmann, B., Hayden, E., Tasche, D., et al. (2003). Measuring the discriminative power of rating systems.

[Fernández-Val, 2009] Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85.

[Firth, 1993] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

[Fuller and Battese, 1974] Fuller, W. A. and Battese, G. E. (1974). Estimation of linear models with crossed-error structure. *Journal of Econometrics*, 2(1):67–78.

[Giancristofaro and Salmaso, 2007] Giancristofaro, R. A. and Salmaso, L. (2007). Model performance analysis and model validation in logistic regression. *Statistica*, 63(2):375–396.

[Greene, 2004a] Greene, W. (2004a). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, 7(1):98–119.

[Greene, 2004b] Greene, W. (2004b). Convenient estimators for the panel probit model: Further results. *Empirical Economics*, 29(1):21–47.

[Greene, 2007] Greene, W. (2007). Discrete Choice Modeling. Working Papers 07-6, New York University, Leonard N. Stern School of Business, Department of Economics.

[Greene, 2012] Greene, W. (2012). *Econometric Analysis, 7th Ed.* Pearson.

[Hahn and Kuersteiner, 2002] Hahn, J. and Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica*, 70(4):1639–1657.

[Hahn and Newey, 2004] Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

[Harrell et al., 1996] Harrell, F. E., Lee, K. L., and Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387.

[Hausman, 1978] Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

[Hoeting et al., 1999] Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401.

[Hsiao, 2003] Hsiao, C. (2003). Analysis of panel data, 2nd. *Cambridge: Cambridge University Press. Kose, Ma, Es Prasad, & Me Terrones (2003), Financial integration and macroeconomic volatility, Imf Staff Papers*, 50:119–142.

[Kosmidis, 2007] Kosmidis, I. (2007). *Bias reduction in exponential family nonlinear models*. PhD thesis, Department of Statistics, University of Warwick.

[Kosmidis and Firth, 2009] Kosmidis, I. and Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika*, 96(4):793–804.

[Kunz et al., 2017] Kunz, J., Staub, K., and Winkelmann, R. (2017). Estimating fixed effects: Perfect prediction and bias in binary response panel models, with an application to the hospital readmissions reduction program.

[Lancaster, 2000] Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of econometrics*, 95(2):391–413.

[Li et al., 2009] Li, D., Bhariok, R., Keenan, S., and Santilli, S. (2009). Validation techniques and performance metrics for loss given default models. *The Journal of Risk Model Validation*, 3(3):3.

[López-Ratón et al., 2014] López-Ratón, M., Rodríguez-Álvarez, M. X., Cadarso-Suárez, C., Gude-Sampedro, F., et al. (2014). Optimalcutpoints: an r package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, 61(8):1–36.

[Lung et al., 2003] Lung, K. R., Gorko, M. A., Llewelyn, J., and Wiggins, N. (2003). Statistical method for the determination of equivalence of automated test procedures. *Journal of Analytical Methods in Chemistry*, 25(6):123–127.

[Maddala, 1986] Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics*. Number 3. Cambridge university press.

[Madigan and Raftery, 1994] Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, 89(428):1535–1546.

[Medema et al., 2009] Medema, L., Koning, R. H., and Lensink, R. (2009). A practical approach to validating a pd model. *Journal of Banking & Finance*, 33(4):701–708.

[Metz, 1978] Metz, C. E. (1978). Basic principles of roc analysis. *Seminars in nuclear medicine*, 8(4):283–298.

[Mundlak, 1978] Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, pages 69–85.

[Neyman and Scott, 1948] Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32.

[on Banking Supervision, 2004] on Banking Supervision, B. C. (2004). *International convergence of capital measurement and capital standards: a revised framework*. Bank for International Settlements.

[Picard and Cook, 1984] Picard, R. R. and Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583.

[Rasero, 2006] Rasero, B. C. (2006). Statistical aspects of setting up a credit rating system.

[Řezáč and Řezáč, 2011] Řezáč, M. and Řezáč, F. (2011). How to measure the quality of credit scoring models. *Finance a úvěr: Czech Journal of Economics and Finance*, 61(5):486–507.

[Steyerberg et al., 2010] Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., and Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, 21(1):128.

[Swamy and Arora, 1972] Swamy, P. and Arora, S. S. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica: Journal of the Econometric Society*, pages 261–275.

[Verbeek, 2004] Verbeek, M. (2004). *A Guide to Modern Econometrics, 2nd Ed.* Wiley.

[Wallace and Hussain, 1969] Wallace, T. D. and Hussain, A. (1969). The use of error components models in combining cross section with time series data. *Econometrica: Journal of the Econometric Society*, pages 55–72.

[Willis, 2006] Willis, J. L. (2006). Magazine prices revisited. *Journal of Applied Econometrics*, 21(3):337–344.

[Wooldridge, 2002] Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. MIT Press.

[Youden, 1950] Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1):32–35.