SNS BANK N.V.

UNIVERSITEIT TWENTE.

Classification: **Extern (Figures censored)**

LOGIT, TOBIT OR HAZARD?

An analysis of modelling the probability default of a retail mortgage owner.

23 September 2015

Master Thesis

Student:
Author: Michiel Verwoerd

Supervisory committee:

First Internal supervisor:	Dr. B. Roorda (Berend)
Second Internal Supervisor:	Drs.ir A.C.M. de Bakker (Toon)
External supervisor:	Dr. S. Muijs (Sandra)
External supervisor:	A. Holst (Annet)

ASN BANK BLG wonen RegioBank SNS Zwitserleven

SNS BANK N.V.   UNIVERSITEIT TWENTE.

Student:

Author: Michiel Verwoerd,

Telephone number: +31 (0)6 xx xx xx xx

Student number: xxxxxxxx

Supervisory committee:

First Internal supervisor:	Dr. B. Roorda (Berend)
Second Internal Supervisor:	Drs.ir A.C.M. de Bakker (Toon)
External supervisor:	Dr. S. Muijs (Sandra)
External supervisor:	MSc. A. Holst (Annet)

Preface

In this report I present my thesis concerning the estimation of 1 year period- and Life Time period estimations of the probability of default of retail mortgage customers. I conducted the research in cooperation with the University of Twente and the Risk Modelling department of SNS Bank N.V.

During my time at SNS Bank N.V. I have gained a lot of experience in the practice of developing risk related models. Now I have got more insight in different underlying frameworks used to predict 'risk'. I'm thankful that I have got the chance to experience how it feels to work for a 'big' professional financial institution. I wish to thank the people who provided me with the successful completion of my internship period.

First of all I would like to thank my external supervisors at SNS Bank N.V.: Annet Holst and Sandra Muijs. Annet and Sandra, thank you for your assistance and all your answers on my questions. I'm thankful for the time you made free for consultation and I enjoyed the Tuesday sessions. I would also like to thank all the other highly skilled people at SNS Bank N.V. that were involved in this research. Because of the involvement of all the colleagues at the SNS Bank N.V. my Thesis was an instructive experience. I can advise everyone who want to learn more about risk modelling to join the Risk Modelling department of SNS Bank N.V. for an internship.

Furthermore, I would like to thank my supervisors at the University of Twente, whom have guided me throughout this research. Berend Roorda, my first supervisor, thank you for your input and suggestions during the meetings in which we discussed the progress of my research. Mr. de Bakker, my second supervisor, thank you for your valuable feedback on my report.

Finally, I want to thank to my loved ones. Thank you for your advice and support.

Michiel Verwoerd

In the train somewhere between Utrecht and Hengelo, August 26, 2015

Management Summary

The goal of this Master's Thesis was to provide the SNS Bank N.V. insight in the 1 year and Life Time (LT) performance (hereafter defined as a period of 5 years) of the Logit-, Tobit- and Hazard framework to determine the Probability of Default (PD) of a retail mortgage customer. This insight is twofold. On one hand SNS Bank N.V. wanted to know which model has the best quantitative performance in the sense of predictive and discriminatory power. On the other hand SNS Bank N.V. wanted to know which framework(s) meets the regulations and the requirements and wishes provided by the model owners and experts. To provide this insight to the SNS Bank N.V., the following research question was set up.

Which framework (Logit, Tobit or Hazard) could SNS Bank N.V. use to estimate the probability of default of a retail mortgage owner, given criteria of predictive and discriminatory power and regulatory requirements?

During this research the theoretical background of the three frameworks and the applicability of the wishes and requirements were studied. The Competing Risk Hazard framework offered a natural way of estimating Life Time PDs, the two other frameworks did not. To be able to predict LT PDs with the Logit and Tobit frameworks, different calculation and extrapolation methods were evaluated. Based on the theoretical background of the frameworks and the methods studied, multiple prototypes were developed (Logit: 1 one year prototype, 3 Life Time prototypes; Tobit: 1 one year prototype; Hazard: 1 multi-period prototype).

During the testing phase of the prototypes, it turned out that the developed Tobit prototype was not able to sufficiently predict PDs, because of a lack of predictive power. This shortcoming was caused by the presence of a level problem as result of a bad fit of the latent Tobit variable. Possible causes of the bad fit could be the high proportion of censored variables, a low predictive power of the variables or a wrong assumption about a not continuously distributed depending variable. The problems with the Tobit model are of such nature that the conclusion has to be drawn that the theoretical idea in combination with the assumptions made in this research, to be able predicting PDs with the Tobit model, did not work.

The best performing prototypes are based on the Logit framework. The 1 year Logit prototype has a good predictive and discriminatory power and therefore meets all the quantitative requirements. All qualitative requirements could be met by selecting variables in consultation with experts, which was not done in this research because of time restrictions. The LT Logit prototype with Least Squares extrapolation is the best LT method regarding predictive and discriminatory power. The results showed that predicting LT PDs is more complex than estimating 1 year PDs, which matches with intuition.

The prototype that is based on the Hazard framework does not meet the requirements regarding predictive power. This is probably caused by the bad fit of the hazard rate of Default to the realizations. This problem could be resolved, in a to be developed prototype. Also improvements in the construction of the dataset and the selection of variables could be made. When these problems are resolved the predictive- and discriminatory power are expected to grow. This outlook and the fact that the Hazard framework offers a natural LT framework in which two events could be fitted, makes the Hazard prototype a promising prototype.

Based on this research the SNS Bank N.V. is advised to use the 1 year Logit prototype and the LT Logit prototype with Least Squares extrapolation (with suggested improvements) as champion model and the Hazard prototype as challenger model. In this setup the Logit prototypes will be used as 'current used' models and the Hazard prototype can be improved as suggested while being live. An overview of the performance of the Logit and Hazard prototypes is presented in paragraph 6.4.

Table of Contents

1 Introduction

In this chapter the context of the research, the research problem itself and the research approach are introduced. This is done by first describing the organizational background in Paragraph 1.1. In Paragraph 1.2 the Master's project description is explained. This is done by first describing the background of the research problem, followed by the description of the core problem and the scope of the research. Paragraph 1.3 presents the methodology, or research approach. This chapter concludes with Paragraph 1.4 in which the outline of this Thesis is given.

1.1 Organizational background

1.1.1 SNS Bank N.V.

The Master's Thesis will be commissioned by the Risk Modelling department of SNS Bank N.V. SNS Bank N.V. is a Dutch retail bank with its headquarter located in Utrecht. SNS Bank N.V. is the 4th of the four major Dutch banks by assets (Nederlandse bankensector) and at this particular moment owned by the Dutch government. The value of the total assets of SNS Bank N.V. in fiscal year 2013 was equal to circa €74 billion of which €53 billion were mortgages (71,6%) (SNS Bank N.V., 2014).

SNS Bank N.V. is a parent company and owns the following brands: ASN Bank, BLG Wonen, Regiobank, SNS bank and Zwitserleven Bank, of which the logos are presented on the title page.

1.1.2 Risk Modelling department

The Risk Modelling department consist of around 10 employees and is a sub-department of the bigger Financial Risk and Modelling (FR&M[1]) department. The task of the Risk Modelling department is to develop and maintain all kinds of risk related models and to support business in quantitative ways. An example of a model developed by the Risk Modelling department is the PHIRM 2.0 PD model (Particuliere Hypotheken Interne Rating Probability of Default Model 2.0), which is used to estimate the Probability of Default (PD) for a retail mortgage customer.

The Risk Modelling department develops models, which are commissioned by model owners. These model owners are the departments who are responsible for the models and/or actually use the models, especially the related estimations generated by the models. These estimates are used for all kind of activities like the determination of regulatory capital, the determination of provisions, pricing of products and much more other activities like general business decisions.

The development of a model is an iterative process. Model owners and experts, with knowledge of the subject to be modelled, are involved during the formulation of the model requirements. The knowledge of the model owners and experts also can be used to determine the factors that will be taken into account in the model.

1.2 Master's project description

1.2.1 Core problem and its background

Currently SNS Bank N.V .is using the Particuliere Hypotheken Interne Rating Probability of Default Model 2.0 (PHIRM 2.0 PD) model to estimate the PD of a retail mortgage customer in the upcoming 12 months (the probability that mortgage costumer X who is not in default at moment of estimation t, goes in default in the period $[t, t+12]$). The development process of this model started in 2013 and the model is released in the

[1] All abbreviations are presented in APPENDIX 1

second half of 2014. The primary goal of the PHIRM 2.0 PD model is to generate PD estimates that are used as input for the determination of Risk Weighted Assets (RWA). The PHIRM 2.0 PD model performs well and meets the requirements set by the model owners and experts (SNS Bank N.V., 2013).

Despite the satisfaction about the PHIRM 2.0 PD model, SNS Bank N.V. is planning to redevelop the model in 2016. The reason for doing this is to meet new regulations and integrate new features. An example of a new requirement that has to be met in the future is the Life Time (LT) requirement (KPMG, 2013), which requires SNS Bank N.V. to be able to estimate LT (time till the end of a contract) PDs, instead of the PD for a period of 12 months.

The process of redeveloping the PHIRM 2.0 PD model into the new Particuliere Hypotheken Interne Rating Probability of Default Model 3.0 (PHIRM 3.0 PD Model) is a radical process, in the sense that the whole model has to be built from scratch. This fact gives SNS Bank N.V. the opportunity to change critical parts of the model, like the underlying framework of the model.

The current underlying framework of the PHIRM 2.0 PD model is the Logit framework (logistic regression framework). This framework is chosen in the past, because it was the industry standard. SNS Bank N.V. wants to investigate if other frameworks can lead to an improvement of the PHIRM 2.0 PD model regarding the quantitative aspect of the estimating performance and the qualitative aspect of the regulations and the requirements and wishes of the model owners and experts. Additionally, an improvement of the model's estimating performance by using another framework, might lead to a decrease in model risk.

1.2.2 Research problem

Following the organisational background and the description of the core problem and its background it becomes clear that SNS Bank N.V. wants to have insight in the performance of alternative frameworks to determine the PD of a retail mortgage customer. This insight is twofold. On one hand SNS Bank N.V. wants to know which model has the best quantitative performance in the sense of predicting power. On the other hand SNS Bank N.V. wants to know which framework meets the regulations and the requirements and wishes provided by the model owners and experts. The regulations and requirements and wishes to be used in this project are regarded as a defined set and are presented in APPENDIX 2 & 3. To solve the core problem by providing SNS Bank N.V. insight in the performance of alternative frameworks, the following research question must be answered.

Which framework (Logit, Tobit or Hazard) could SNS Bank N.V. use to estimate the probability of default of a retail mortgage owner, given criteria of predictive and discriminatory power and regulatory requirements?

Before concluding on the main question, several sub-questions need to be answered. The first sub-question that needs to be answered is stated below.

1. **What is a Logit framework and how is it used to estimate the probability of default (PD) of a retail mortgage customer?**

The first sub-question is meant to give insight in the current framework used. This insight will help understanding the PHIRM PD 2.0 model, which is the model that needs to be adapted. The Logit framework will also be used as benchmark.
The second and third sub-question that need to be answered are provided below.

2. **What is a Tobit framework and how could it be used to estimate the probability of default (PD) of a retail mortgage customer?**

3. **What is a Hazard framework and how could it be used to estimate the probability of default (PD) of a retail mortgage customer?**

The second and third sub-question are meant to give insight in the alternative frameworks. This insight is needed before estimating the probability of default with those frameworks. The 'how part' of the question is meant to gain insight how the probability of default could be estimated with those frameworks.

The fourth sub-question that needs to be answered is mentioned below.

4. **How can the regulations and the requirements and wishes given by the model owner and experts, which have a significant impact on the design of the prototypes, be met?**

The fourth sub-question is meant to find out how the regulations and the requirements and wishes can be met during the application of each framework. The regulations and the requirements and wishes to be used in this project are presented in Appendix 3 & 4. The life time requirement as described in Paragraph 1.1.2 will become one of the main issues during this research. For this, but also for the other requirements, holds that their existence is not a subject to this study. This sub question, and more in general this Thesis, focusses on how the regulations and requirements and wishes can be met.

The fifth sub-question that needs to be answered is the following.

5. **How to test the performance of the different frameworks regarding the quantitative aspect of PD estimating and the qualitative aspect regarding the regulations and the requirements and wishes of the model owners and experts?**

The fifth sub-question is meant to give insight in how to compare the performance of the 3 frameworks. All frameworks have to be compared in the same way to provide a thought-out conclusion on the main question.

1.2.3 Scope

Many frameworks could be used for estimating the probability of default. For this research the decision is made to evaluate the Tobit framework and the Hazard framework. This choice is made because SNS Bank N.V. has the highest expectations regarding the performance of these two frameworks. The Logit framework will be used as benchmark, this framework has proven its existence in the current model and will only be replaced if one of the other frameworks will outperform the Logit framework.

Besides the restriction regarding the models to evaluate, some other restrictions are set out to keep a clear focus on SNS Bank N.V.'s main questions to be addressed. In consultation with the external supervisors is decided that bucketing is outside the scope of this research. As result of this decision real PDs will be compared with realized rates, which facilitate a fair comparison of the models quantitative performance.

Also decided is that the focus in this thesis is on customers that go into default after being in arrears. This because the behaviour of the group of customers that goes into default after they sold their house for less than their mortgage is different from the 'regular' defaulters.

Another decision made is that static versus dynamic modelling (also known as Point in Time versus Trough The Cycle modelling) is outside scope. One of the main reasons for this decision is that there is no data of sufficient quality available in the datamart. Other decisions that limit the scope of this project, like the default definition and which regulations to meet, are stated in the overview of regulations and requirements and wishes to be used in this project presented in Appendix 2 & 3.

Furthermore is decided that the focus will be on 1 year and 5 year PD estimations. The 1 year PD estimations are of importance according Basel regulations. The 5 year PD estimations represent life time PD estimations. The choice to represent a life time period as 5 years is data related. The dataset used contains 5 years of data that could be used for testing purposes.

1.3 Methodology

The process of evaluating the different frameworks and picking the best alternative is illustrated in Figure 1-1.

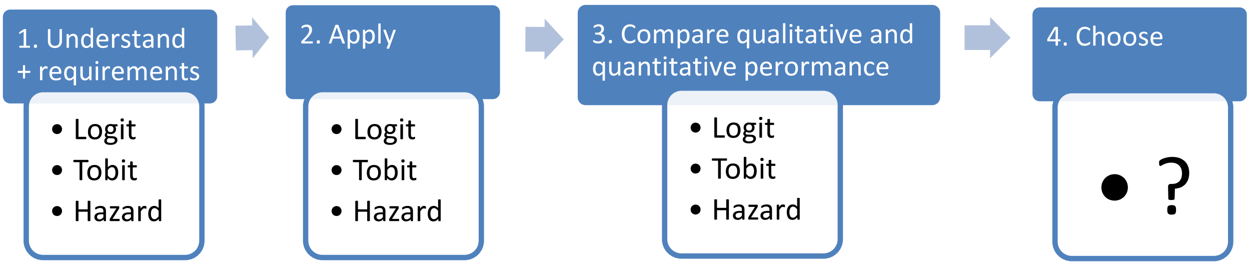


Figure 1-1 The process of evaluating the different frameworks and picking the best alternative

As can be observed in Figure 1 the first step is related to the first 4 sub-questions and the applicability of the requirements. In this step all information regarding the frameworks will be gathered. Literature will be studied to understand the different frameworks and their practical application. In the first step also the regulations and the requirements and wishes given are studied. For some of the requirements this study will be quite straightforward. An example of such requirement is that the variables that may be used have to be part of the short list of SNS Bank N.V. or have been used in the current PHIRM 2.0 PD model. For other requirements the applicability study will be more in depth. An example of such a requirement is the life time requirement. For this requirement it is not clear without further study how to meet it.

In the second step of the process the different frameworks will be applied. A prototype model will be built in Matlab for each framework. In this step the information gathered in the first step about how to apply the frameworks will be used.

The third step is closely related to sub-question 5. In this step the answer on sub-question 5, how to measure the qualitative and quantitative performance of the different frameworks, will be given. Literature will be used to provide an answer to this question. This answer will be used in the same step to compare the performance of the alternatives.

In the fourth and final step, the best performing framework given the criteria and requirements will be chosen. A recommendation will be formulated in which is stated which framework SNS Bank N.V. could use to estimate the PD of a retail mortgage customer.

1.4 Outline

Chapter 2 of this Thesis contains the Theoretical framework. In this chapter all theory used in this Thesis is explained. This theory comprises the frameworks studied to answer the first 3 sub-questions and also the methods which are studied to answer the 5th sub-question about measuring the performance of the different prototypes. The description of the frameworks has to be interpreted as the theoretical idea about how the different frameworks could be used to predict PDs. During the research it turned out that not all theoretical ideas worked out as well as intended. Problems that occurred during the practical application of the frameworks are analysed and provided with a direction of solution or a direction of improvement. The results of this effort are presented in Chapter 6.

Chapter 3 of this Thesis contains the answer on sub-question 4 and gives insight in how the more complicated requirements and wishes of model owners and experts can be met. Straightforward requirements and wishes like will not be treated in this chapter. The more straightforward requirements and wishes will be referred to in chapters where these requirements and wishes are relevant as part of the assessment of qualitative performance in chapter 6.

Chapter 4 contains a description of all the developed prototypes and sub prototypes. In this chapter the process of variable selection is described for each prototype and in addition the variables which are selected for each prototype are presented.

Chapter 5 contains a short description about the used data and in what way the data is used.

Chapter 6 contains the results of estimating PDs with the developed prototypes and an assessment of performance, based on the methods according the 5th sub-question. For frameworks of which the practical application did not turn out as theoretically intended, the problems that occurred are analysed and provided with a direction of solution or a direction of improvement.

Chapter 7 presents the conclusion and the recommendations. A summary is given followed by a conclusion and advice towards the SNS Bank N.V.

Chapter 8 contains the discussion and presents suggestions for further research.

2 Theoretical Framework

In this chapter all theoretical background used in this Thesis is explained. It contains a description of the different frameworks and the way how the frameworks will be used to estimate PDs is evaluated. Also the methods that are used to measure the performance of the prototypes are described.

2.1 Logit Framework

The Logit Framework, or logistic regression framework, is a direct probability framework that can handle binary outcomes (Cox, 1958). The Logit function, described in Equation 2.1, defines the relation between the K independent variables $(x_1, \dots, x_k)$, their weights $(\beta_0,\ \beta_1,\ \dots,\ \beta_n)$ and the dependent variable (PD), which is the probability that a certain binary event will happen.

$$PD = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_k)}} \tag{2-1}$$

In the context of this Research the binary event is defined as at least one default event[2] of a retail mortgage customer within a defined time range $[\mathrm{t}=0, \mathrm{t}+\mathrm{m}]$. For the purpose of understanding the Logit function is illustrated in Figure 2-1 (generated with synthetic data). The figure shows the binary outcome per customer and the Risk indicator score $(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_k)$.

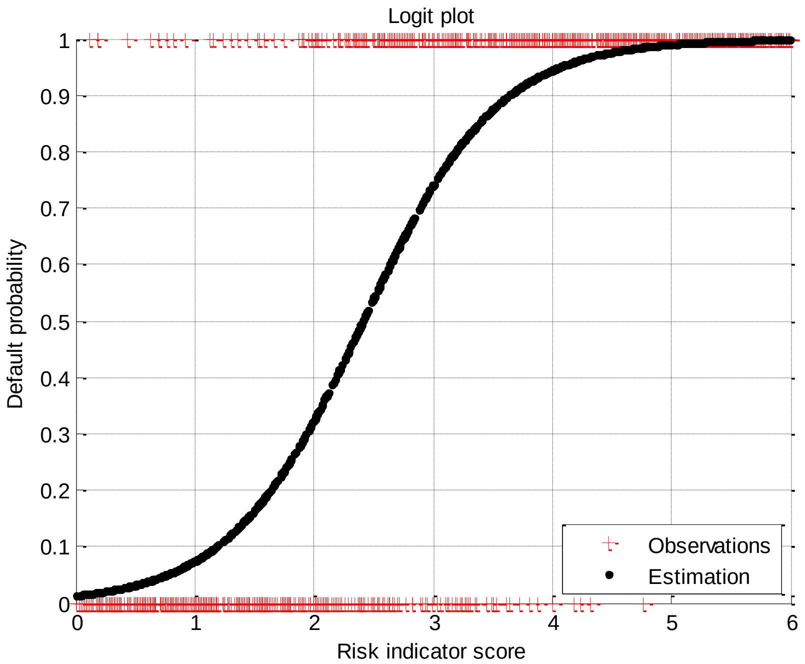


Figure 2-1 The Logit Function

The Logit framework can be used to estimate PDs by first determining the relation between realizations and risk indicators, based on historical data ($N\ observations$). In other words, estimate the weights ($\beta_0,\ \beta_1,\ \dots,\ \beta_k$), using the Maximum Likelihood Function described in Equation 2.2 (Dobson, 2002).

$$L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n) = \prod_{i=1}^{N} \left(\frac{1}{1+\ e^{-(\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_n x_k)}}\right)^{y_i} * \left(1 - \frac{1}{1+\ e^{-(\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_n x_k)}}\right)^{1-y_i} \tag{2-2}$$

$$with\ y_i = \begin{cases} 0 & (not\ in\ default) \\ 1 & (in\ default) \end{cases}$$

[2] $Default\ definition$: (ammount arrears / monthly paymentammount $\geq$ 3) $= Default$

An important assumption is that the relation described in Equation 2.1, did hold in the past and will also hold in the future. When this assumption is made, one can predict the probability of default of a retail mortgage customer given any set of risk indicators $(x_1, \dots, x_k)$.

Important to notice is that the application of the Logit framework in this Thesis defines the relation between a Risk indicator score and the probability of an event happening in a fixed time range. Given the fact that the life time in reality differs per customer and it is not supposed to develop different prototypes for all these life times, an extrapolation method has to be used to be able of estimating life time PDs per customer. The extrapolating techniques used are presented in Chapter 3.

2.2 Tobit Framework

The Tobit framework can handle censored (a.k.a. limited) data (Woolridge, 2002). Because of this feature the Tobit framework can define the relation between K independent variables $(x_1, \dots, x_k)$, their weights (β_0, β_1, … , β_n) and a continuous limited depending variable (y_i) (Tobin, 1958).

In the application of the Tobit Framework, the limited depending variable (y_i), also referred to as 'default score', is assumed to be continuous and equal to $(\text{ammount of arrears / monthly payment ammount})$. This depending variable has a lower limit, equal to 0, because the dataset available does not support a negative amount of arrears and does not contain negative 'monthly payment amounts'. Instead of a negative amount of arrears, a zero is presented in the dataset, which makes the default score for observations with a amount of arrears equal to zero, also equal to zero $(0 \text{ / monthly payment ammount} = 0)$.

The Tobit framework relation between K independent variables $(x_1, \dots, x_k)$, their weights (β_0, β_1, … , β_n) and the observed default score (y_i) is described by the latent variable (y_i^*), as described in Equation 2-3. The difference between the observed default score (y_i) and the latent variable (y_i^*), is that the observed default score has a lower limit, equal to zero, and the latent variable is unlimited. Therefore, the latent variable (y_i^*) represents the relation between the Risk indicator score $(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_k)$ and the default score, for both the observable $(y_i^* > 0)$ range and unobservable $(y_i^* \leq 0)$ range.

$$Observable\ y_i \quad = \begin{cases} y_i^* & if\ y_i^* > 0 \\ 0 & if\ y_i^* \leq 0 \end{cases} \qquad with\ latent\ variable\ y_i^* = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_k \tag{2-3}$$

In Figure 2-2, which is generated with synthetic data for purpose of illustration, the relation between the Risk indicator score $(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_n x_k)$, the observations $(\text{amount of arrears / monthly payment amount})$ and the latent variable (y_i^*) is illustrated. The observations (red) are the observations as presented in the dataset used. The observations (red) which have a default score equal to 0, are the censored observations, as described in Equation 2-3. One can assume that these censored observations are representations of observations with hidden negative default scores (blue). The estimated latent variable y_i^* (black) represents the relation between the Risk indicator score and the default score, for both the observable $(y_i^* > 0)$ and unobservable $(y_i^* \leq 0)$ range.

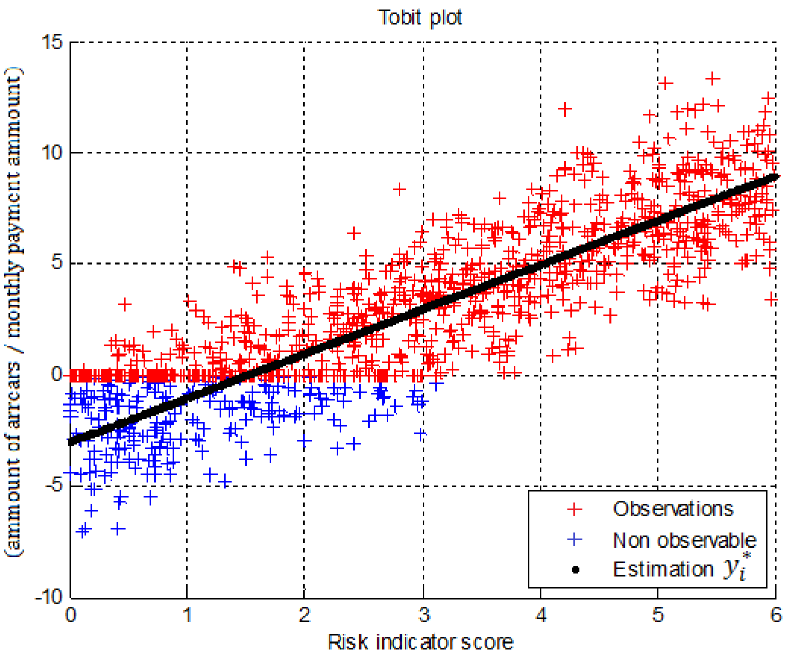


Figure 2-2 The Tobit Function

Because the Tobit framework has, in contrast to the Logit framework, a score as dependent variable instead of a probability, an extra transformation is needed to calculate PDs from the estimated y_i^*. This transformation can be made by using the assumption that the Tobit error terms in this Thesis ($u_i =$ the gaps between y_i^* and y_i) are assumed to follow a Normal distribution ($N(0, \sigma^2)$). Which normally would be an result of a good fit of y_i^* (Tobin, 1958). Because of the assumed Normal distribution a σ can be estimated based on all the error terms. This σ, together with the μ, which is equal to y_i^*, can be used in the cumulative distribution function of the Normal distribution to calculate the PDs, as defined in Equation 2.4. Notice that the default definition implies that a customer is in default as $(\text{arrears / monthly payment} \geq 3)$, where $(\text{arrears / monthly payment})$ is equal to the default score used in the Tobit model.

$$PD_i = 1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{3 - y_i^*}{\sigma\sqrt{2}}\right)\right], \qquad \textbf{(2-4)}$$

$y_i^* = estimated\ latent\ Tobit\ variable$

$\sigma = standard\ deviation\ of\ the\ normally\ distributed\ error\ term\ u_i$

The weights and sigma ($\beta_0,\ \beta_1,\ \dots,\ \beta_n, \sigma$) of the Tobit model can be estimated with the help of the Maximum Likelihood Function, described in Equation 2-5 (Amemiya, 1973). Notice that handling the censored (a.k.a. limited) observations is done in the formula by the $I(y_i)$ indicator.

$$L(\hat\beta_0, \hat\beta_1, \dots, \hat\beta_n, \hat\sigma) =$$

$$\prod_{i=1}^{N}\left(\frac{1}{\hat\sigma}\phi\left(\frac{y_i - (\hat\beta_0 + \hat\beta_1 x_1 + \dots + \hat\beta_n x_{k)})}{\hat\sigma}\right)\right)^{I(y_i)}\left(1 - \Phi\left(\frac{(\hat\beta_0 + \hat\beta_1 x_1 + \dots + \hat\beta_n x_{k)} - y_L}{\hat\sigma}\right)\right)^{1-I(y_i)}, \qquad \textbf{(2-5)}$$

$with\ I(y_i) = \begin{cases} 0 & if\ y_i = 0 \\ 1 & if\ y_i > 0 \end{cases}$

$\phi =\ the\ PDF\ of\ the\ normal\ distribution$

$and\ \ \Phi =\ the\ CDF\ of\ the\ normal\ distribution$

$y_L = 0\ (censoring\ boundary)$

An important assumption is that the relation described in Equation 2.3, did hold in the past and will also hold in the future. When this assumption is made, one can predict the probability of default of a retail mortgage customer given any set of risk indicators $(x_1, \dots, x_k)$.

The Tobit framework is just as the Logit framework a framework that defines the relation between the probability of an event happening in a fixed time range and a Risk indicator score. For the same reasons as for the Logit framework extrapolation is needed. The extrapolation techniques that will be used are the same as for the Logit framework.

2.3 Hazard Framework

The Hazard framework is, in contrast to the Logit- and Tobit framework, a multi-period framework that defines the relation between the Risk indicator and the PD for all time periods between the start of a customer's contract and the life time of a customer's contract. Therefore one could argue that the Hazard framework is more appropriate for forecasting defaults (Shumway, 1999). In this Thesis a competing risk hazard framework is used (B. A. Ciochetti, 2002), which is illustrated in Figure 2-3. The reason for using this framework is the limited variation of events that can happen during a customer's contract period. Customers will either go in default, make a total prepayment (pay all their outstanding mortgage debt) or stay healthy. In this situation prepayment is the competing risk of default.

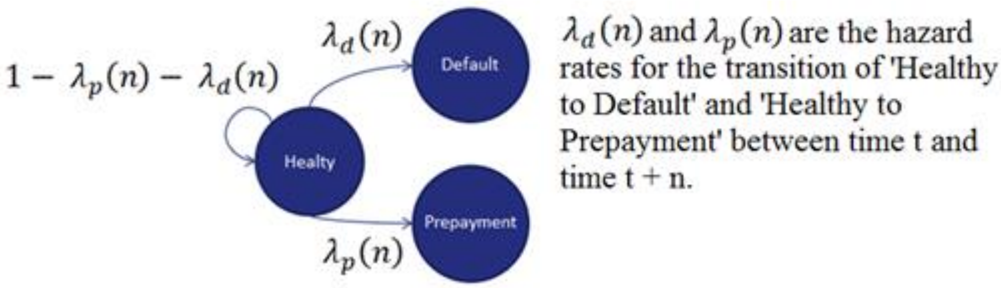


Figure 2-3 Competing Risk Framework

The relation between the hazard rates and the PD in the next t months per customer is in the competing risk framework defined by the Cumulative Incidence Function of default (CIF), given in Equation 2-6 (Rodríguez, Competing Risks, 2005). The CIF for default represents the cumulative probability of default after t months. The default curve is a sub distribution, because the probability does not approach 1 when $t \to \infty$, as result of the competing risk of total prepayment. T is defined as loan age at exit event, where the exit events could be default and prepayment.

$$PD(t) = \mathbb{P}(T \leq t \cap D = 1) = \int_0^t \lambda_{def}(\tilde{t})\, S(\tilde{t}) d\tilde{t}, \quad \textbf{(2-6)}$$
$$with\ t = time\ in\ months, T = loan\ age\ at\ exit\ event\ and\ D = \begin{cases} 0\ (not\ in\ default) \\ 1\ (in\ default) \end{cases}$$

λ_{def} is the hazard rate for the transition from healthy to default, defined as: **(2-7)**
$$\lambda_{def}(t)dt \equiv \mathbb{P}(t \leq T < t + dt, D = 1)|T \geq t)$$

$S(t)$ is the 'overall' survival curve (given the competing risks of default and prepayment): **(2-8)**
$$S(t) = \mathbb{P}(T > t) = \exp\big(-\Lambda(t)\big),$$
$$\text{with } \Lambda(t) = \int_0^t \lambda(\tilde{t}) d\tilde{t}\ \ and$$

$$\Lambda(t) = \Lambda_{def}(t) + \Lambda_{prep}(t) = \int_0^t \lambda_{def}(\tilde{t})\, d\tilde{t} + \int_0^t \lambda_{prep}(\tilde{t})\, d\tilde{t}$$

λ_p is the hazard rate for the transition from healthy to total prepayment and therefore it holds that $\lambda(t) = \lambda_{def}(t) + \lambda_{prep}(t)$ (the total hazard rate λ equals the hazard rate for the two end state), as illustrated in figure 2.3.

The framework has to be able to estimate on a monthly basis and therefore also has to cope with customers with a $loan\ age > 0$. This dynamic aspect is accomplished by determine the CIF at the age of a loan ($t_a = loan\ age\ in\ months\ at\ moment\ of\ measurement$), as described in Equation 2-9.

$$PD(t|T > t_a) = \int_{t_a}^{t} \lambda_{def}(\tilde{t})\, S(\tilde{t}|T > t_a) d\tilde{t} = \int_{t_a}^{t} \lambda_{def}(\tilde{t}) \frac{S(\tilde{t})}{S(t_a)} d\tilde{t} \qquad \textbf{(2-9)}$$

This follows from, using Bayes theorem:

$$S(t|T > t_a) = \mathbb{P}(T > t|T > t_a) = \frac{\mathbb{P}(T>t \cap T>t_a)}{\mathbb{P}(T>t_a)} = \frac{\mathbb{P}(T>t)}{\mathbb{P}(T>t_a)} = \frac{S(t)}{S(t_a)}, if\ t > t_a \qquad \textbf{(2-10)}$$

The probability distributions of Defaults and Prepayments, and thereby also their hazard rates, are assumed to be Weibull distributed. This assumption is based on two reasons. The first reason is that the Weibull distribution is one of the most useful distributions regarding Life Time data (C.B. Guure, 2002). The second reason is that plots per cohort of realized default and prepayment frequencies showed evidence that the probability of both events is Weibull distributed. The Weibull hazard rate for default and the Weibull survival functions, presented in Equations 2-11, 2-12 and 2-13.

Weibull hazard rate for default: $$\lambda_{def} = \left(\frac{\alpha_{def}}{\tau_{def}}\right)\left(\frac{\tilde{t}}{\tau_{def}}\right)^{\alpha_{def}-1} \qquad \textbf{(2-11)}$$

Weibull survival function for default and prepayment: $$S(\tilde{t}) = -\left(\left(\frac{\tilde{t}}{\tau_{def}}\right)^{\alpha_{def}} + \left(\frac{\tilde{t}}{\tau_{prep}}\right)^{\alpha_{prep}}\right) \qquad \textbf{(2-12)}$$

Weibull survival for age: $$S(t_a) = e^{-\left(\left(\frac{t_a}{\tau_{def}}\right)^{\alpha_{def}} + \left(\frac{t_a}{\tau_{prep}}\right)^{\alpha_{prep}}\right)} \qquad \textbf{(2-13)}$$

The CIF used is based on this Weibull assumption and is presented in Equation 2-14. Equation 2-14 is the same as equation 2-9, complemented with the Weibull hazard rate for default and the Weibull survival functions, presented in Equation 2-11, 2-12 and 2-13.

$$PD(t|T > t_a) = \int_{t_a}^{t} \lambda_{def}(\tilde{t}) \frac{S(\tilde{t})}{S(t_a)} dt = \int_{t_a}^{t} \left(\frac{\alpha_{def}}{\tau_{def}}\right)\left(\frac{\tilde{t}}{\tau_{def}}\right)^{\alpha_{def}-1} * \frac{e^{-\left(\left(\frac{\tilde{t}}{\tau_{def}}\right)^{\alpha_{def}} + \left(\frac{\tilde{t}}{\tau_{prep}}\right)^{\alpha_{prep}}\right)}}{e^{-\left(\left(\frac{t_a}{\tau_{def}}\right)^{\alpha_{def}} + \left(\frac{t_a}{\tau_{prep}}\right)^{\alpha_{prep}}\right)}} * d\tilde{t}, \qquad \textbf{(2-14)}$$

$with\ \tau_{def} = (\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + u_i),$
$u_i \sim N(0, \sigma^2),$
$\alpha_{def} = the\ scale\ parameter\ for\ default,$
$\tau_{prep} = the\ shape\ parameter\ for\ total\ prepayment\ and$
$\alpha_{prep} = the\ scale\ parameter\ for\ total\ prepayment$

The scale parameter of the Weibull distribution for defaults, the τ_d parameter, used in Equation 2.6., is the only parametric parameter in this equation. The reason for making the scale parameter dependent is based on the fact that the scale parameter is the most dominant parameter in shaping the distribution. Making both parameters, shape and scale, parametric was not desirable because it complicates the maximum likelihood estimation and thereby raises the chance of errors. As an in depth study of prepayment is outside scope the parameters of the Weibull distribution for prepayment are chosen to be constant.

The Hazard competing risk framework can be used to estimate PDs by first determining the relation between realizations, risk indicators and other parameters, based on historical data. In other words, estimate the Tau's and Alpha's ($\hat{\tau}_{def} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_n, \hat{\alpha}_{def}, \hat{\tau}_{prep}, \hat{\alpha}_{prep}$). This is done with the help of two Maximum Likelihood Functions, one for default and one for prepayment. Both Maximum Likelihood Functions are distracted from the Maximum Likelihood Function for censored data, described in equation 2-15 (Zhang, 2005).

$$L(\theta, x, \delta) = \prod_{i=1}^{N} [f(x_i;\theta)]^{\delta_i} * [F(x_i;\theta)]^{1-\delta_i} \qquad \textbf{(2-15)}$$

$$with, \theta\ are\ the\ parameters\ of\ interest, x\ is\ the\ data\ and\ \delta_i\ is\ the\ censored\ indicator,$$
$$f(x_i;\theta) = the\ pdf\ of\ the\ assumed\ distribution, F(x_i;\theta) = the\ cdf\ of\ the\ assumed\ distribution$$

The Maximum Likelihood Function for default is presented in Equation 2-16, it contains the PDF and CDF of the Weibull distribution corrected for the loan age at the start of the observation period, by dividing both probability functions by the probability of survival. Notice that the loan age at the start of the observation period (a_i), used during estimating the prototype, is not the same as the loan age at the moment of measurement (t), which is used for calculation of the PDs. Furthermore can be observed in the equation that correct handling of censored data, by ($c_{i,def}$), is of great importance (J. D. Kalbfleisch, 2002). When a default event happens in the period of observation, the observation is censored for prepayment and vice versa. When nothing happens during the period of observation, the observation is censored for both default and prepayment[3].

$$L(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n, \hat{\alpha}_{def}) = \prod_{i=1}^{N} \left[\frac{\left(\left(\frac{\hat{\alpha}_{def}}{\hat{\tau}_{i,def}}\right) * \left(\frac{t_i}{\hat{\tau}_{i,def}}\right)^{\hat{\alpha}_{def}-1} * e^{-\left(\frac{t_i}{\hat{\tau}_{i,def}}\right)^{\hat{\alpha}_{def}}} \right)}{e^{-\left(\frac{a_i}{\hat{\tau}_{i,def}}\right)^{\hat{\alpha}_{def}}}} \right]^{c_{i,def}} * \left[\frac{e^{-\left(\frac{t_i}{\hat{\tau}_{i,def}}\right)^{\hat{\alpha}_{def}}}}{e^{-\left(\frac{a_i}{\hat{\tau}_{i,def}}\right)^{\hat{\alpha}_{def}}}} \right]^{1-c_{i,def}}, \qquad \textbf{(2-16)}$$

$$with\ i = customer,$$
$$\hat{\tau}_{i,def} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_n,$$
$$c_{i,def} = \begin{cases} 1 & (customer\ i\ defaulted) \\ 0 & (customer\ i\ did\ not\ defaulted) \end{cases},$$
$$t_i = time\ between\ start\ observation\ and\ event\ and$$
$$a_i = loan\ age\ in\ months\ at\ the\ start\ of\ the\ observation$$
$$N = number\ of\ observations$$

[3] More information about the correct handling of censored data is given in chapter 5.

The Maximum Likelihood Function for prepayment, described in Equation 2-17, is pretty similar to the one for default. Also for this function holds that the correct handling of censored observations is of big importance.

$$L(\hat{\tau}_{prep}, \hat{\alpha}_{prep}) = \prod_{i=1}^{N} \left[\frac{\left(\left(\frac{\hat{\alpha}_{prep}}{\hat{\tau}_{prep}}\right) * \left(\frac{t_i}{\hat{\tau}_{prep}}\right)^{\hat{\alpha}_{prep}-1} * e^{-\left(\frac{t_i}{\hat{\tau}_{prep}}\right)^{\hat{\alpha}_{prep}}} \right)}{e^{-\left(\frac{a_i}{\hat{\tau}_{prep}}\right)^{\hat{\alpha}_{prep}}}} \right]^{c_{i,prep}} * \left[\frac{e^{-\left(\frac{t_i}{\hat{\tau}_{prep}}\right)^{\hat{\alpha}_{prep}}}}{e^{-\left(\frac{a_i}{\hat{\tau}_{prep}}\right)^{\hat{\alpha}_{prep}}}} \right]^{1-c_{i,prep}}, \quad \textbf{(2-17)}$$

$$\begin{aligned}
&with\ i = customer,\\
&c_{i,prep} = \begin{cases} 1 & (customer\ i\ did\ a\ total\ prepayment) \\ 0 & (customer\ i\ did\ not\ a\ total\ prepayment)' \end{cases}\\
&t_i = time\ between\ start\ observation\ and\ event\ and\\
&a_i = \ loan\ age\ in\ months\ at\ the\ start\ of\ the\ observation\\
&N = number\ of\ observations
\end{aligned}$$

To be able to predict PDs with the Hazard framework, the assumption has to be made that the relation described in Equation 2.6, did hold in the past and will also hold in the future. When this assumption is made, one can estimate the $PD(t|T > t_a)$ of a retail mortgage customer given any set of risk indicators $(\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n})$.

2.4 Performance Measuring

2.4.1 Quantitative performance

Quantitative performance can be divided in two categories, namely discriminatory power and predictive power. The first category gives information about how well a framework performs regarding arranging of customers. A framework with high discriminatory power is better in discriminating defaulters from non-defaulters (Wu, 2008). A framework with high predictive power performs well regarding the level of predictions. In this Thesis the performance of the frameworks on both categories are tested with multiple tests.

Discriminatory power

To assess the discriminatory power of the frameworks the Area Under the Receiver Operating Characteristic curve (AUROC) and the Kolmogorov-Smirnov methods are used. Especially the first method is commonly used and promising since the statistical properties are well investigated (Tasche, 2006). Both methods are shortly described.

<u>The Area Under the Receiver Operating Characteristic curve (AUROC)</u>

The ROC is a curve which represents the discriminatory power of a binary classification system (Wu, 2008). To create the curve, all observations have to be labelled as being 'positives' or 'negatives'. The first category, 'the positives', represents the customers who defaulted. The second category, 'the negatives', represents the customers who didn't go into default. After the labelling, all observations have to be ranked in a descending order based on their estimated PD (on time $[\mathrm{t}-\mathrm{n},]$). Then the curve can be constructed, starting at $(0,0)$. For each 'positive' observation a step $(1\ /\ \#\ of\ 'positive'observations\ )$ up has to be made, for each 'negative' observation a step $(1\ /\ \#\ of\ 'negative'observations\ )$ to the right has to be made. The result of this process, the ROC curve, is illustrated as an example in Figure 2-4. The closer the curve comes to the left upper corner, the better the discriminatory performance of the model is. If the curve is diagonal, the model estimations are random. The performance described above can be expressed as an AUROC, the area under the ROC curve, of which the interpretation which is commonly used is given in Table 2-1.

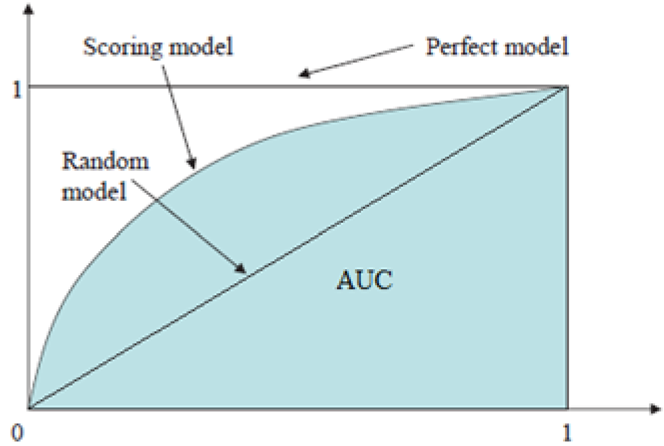


Figure 2-4 The ROC and AUC, retrieved from (Wu, 2008)

AUROC	Result
50-60%	weak
60-70%	moderate
70-80%	sufficient
80-90%	good
90-100%	excellent

Table 2-1 AUC scores and their interpretations, retrieved from (Swets, 1988)

<u>The Kolmogorov-Smirnov test</u>
The Kolmogorov-Smirnov test is a test to assess if two data sets have the same or a different underlying distribution (G. Marsaglia, 2003). In this Thesis the test will be used to assess if the distribution of PDs of defaulted customers significantly differ from PDs of non-defaulted customers. The Kolmogorov-Smirnov test is an appropriate test for this purpose, because the test can be performed without any assumption about parameters or a distribution. The test will be conducted with a significance level of $\alpha = 0.05$.

Predictive power
To assess the predictive power of the frameworks the Binomial test, the Normal test and the Vasicek test are used. In this Thesis static versus dynamic modelling is out of scope, therefore it could be possible that some or all tests advise to reject the hypotheses. This is due to the fact that all the tests are dynamic or static orientated and the frameworks don't generate purely static or dynamic estimates. So the results of the tests for predictive power have to be used as an indication. All tests are shortly described. They will be conducted with a significance level of $\alpha = 0.05$.

<u>The Binomial test</u>
The Binomial test uses the binomial distribution to test if binary realizations (Default Rate (DR)) fall into the confidence interval of the estimates (Howell, 2007). To use the Binomial test all the observations will be divided into buckets with the bucket borders which are used in the PHIRM 2.0 PD model[4]. Then per bucket a confidence interval will be created and per moment of measurement will be checked if the DR falls into the confidence interval. As the Binomial test is clearly a Point in Time (PIT) test, the percentage of DR's that would not fit into a PD interval is expected to be higher than 5%.

<u>The Normal test</u>
The Normal test is in contrast to the other tests for predictive power a typical TTC test (Tasche, 2006). The test estimates per bucket the possible deviation from the long term average by looking per moment of

[4] APPENDIX 5

measurement $(peil\_dt)$ at the realized deviations. The assumption is made that the deviations tested are normally distributed. Then the test checks if the realized deviations deviate significantly from the estimates.

The Vasicek test

The Vasicek test is a test which does not assume observations to be uncorrelated and accounts for the worst economic influences in the credit cycle (C. Huber-Carol, 2002). This is in contrast with the earlier described tests, which assumes that the observations per moment of measurement $(peil\_dt)$ are not correlated. This assumption of non-correlation is unlikely to hold in reality, because more customers go into default in bad economic times. A significance level of $\alpha = 0.05$ implies that the worst economic influences in a cycle of 20 years are taken into account. All DR's per bucket are compared with the calculated Vasicek PDs per moment of measurement $(peil\_dt)$. The Vasicek test is a TTC test, but since the Vasicek PD is much higher than the estimated PD the percentage of DR's that do not fall into a PD interval is expected to be lower than 5%.

2.4.2 Qualitative performance

The qualitative performance, or the degree to which the different frameworks meet the qualitative requirements and serve the wishes of the model owners, will be assessed by one of the external supervisors. The external supervisor will assess if the quality of the prototypes, based on the different frameworks, is of sufficient level to deliver the prototypes to the model owners.

3 Requirements and Wishes

In this chapter the approach to meet the more complicated requirements and wishes of the model owners and experts is explained. Straightforward requirements and wishes like '*The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.*', are out of scope in this chapter. To the more straightforward requirements and wishes will be referred in chapters where these requirements and wishes are relevant or during the assessment of performance described in chapter 6.

3.1 Life time

The Life Time requirement, formulated as: *The models that will be developed during this research have to be able to estimate Life Time (LT) PDs. LT is defined as: "The period between the start and the end of a customer's contract."* is a requirement that follows from the new IFRS 9 regulation (KPMG, 2013). In this new regulatory framework is prescribed that banks need to use life time PDs during the calculation of provisions for customers of which the creditworthiness is significantly decreased.

As explained in chapter 2, the Hazard framework offers with the Cumulative Incidence Function a natural way of calculating life time PDs. The Logit and Tobit framework do not offer a natural way of calculating Life Time PDs. To be able to predict life time PDs with the latter two frameworks three extrapolation methods will be evaluated. An example of extrapolation is illustrated in Figure 3-1. In this Thesis the Logit and Tobit PDs will be estimated for $year\ 0, 1, 2\ and\ 3$ and with the help of extrapolation the LT PD $5\ year$ will be calculated.

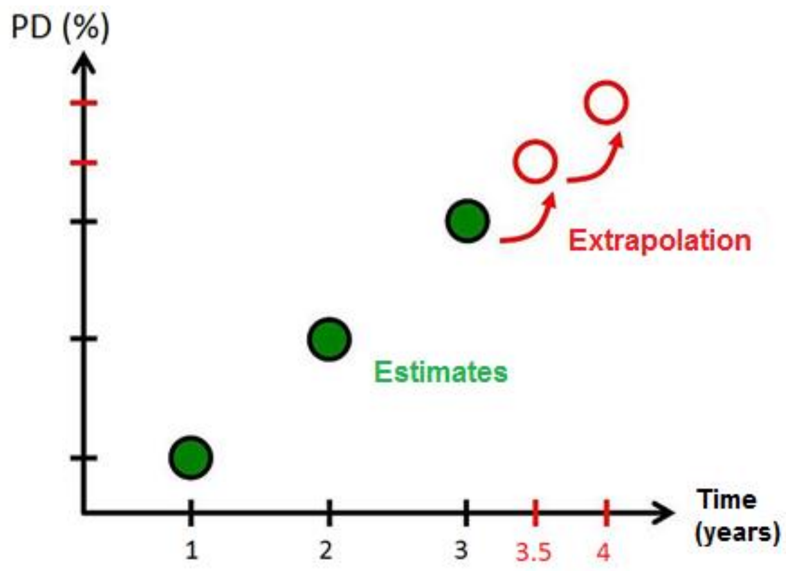


Figure 3-1 Extrapolation example

3.1.1 Extrapolation – mean Lambda

By applying the mean Lambda extrapolation method, the assumption is made that PDs are cumulative Exponential distributed as described in Equation 3-1. The first step of the method is to extract the Lambda for each unique estimation (PD_i). The second step is to calculate the mean Lambda on a customer level. In the last step the mean Lambda of a unique customer is used to calculate the LT PD. An example of the mean Lambda extrapolation method is presented for a fictitious customer on the next page.

$$PD_{it} = 1 - e^{-\lambda_i t}, \qquad with\ i = customer\ and\ t = period\ (in\ years) \tag{3-1}$$

Extrapolation – mean Lambda example

The PD estimations corresponding with the different time periods, of customer X, are presented in Table 3-1.

3-1 PD estimations for fictitious customer X per time period

Time period (years)	PD estimation (%)
1	8.39%
2	14.72%
3	19.15%

Step 1 (extract Lambda for observations of fictitious *customer X):*

$$\lambda_1 = -\frac{\ln(1 - PD_{x1})}{1} = -\frac{\ln(1 - 0.0839)}{1} = 0.0876$$

$$\lambda_2 = -\frac{\ln(1 - PD_{x2})}{1} = -\frac{\ln(1 - 0.1472)}{2} = 0.0796$$

$$\lambda_3 = -\frac{\ln(1 - PD_{x3})}{1} = -\frac{\ln(1 - 0.1915)}{3} = 0.0709$$

Step 2 (calculate mean Lambda for fictitious *customer X):*

$$\bar{\lambda} = \frac{(\lambda_1 + \lambda_2 + \lambda_3}{3} = \frac{0.0876 + 0.0796 + 0.0709}{3} = 0.0793$$

Step 3 (calculate LT PD for fictitious *customer X):*

$$PD_{xLT} = 1 - e^{-\bar{\lambda}_x LT} = 1 - e^{-0.0793 * 5} = 0.3273 = 32.73\%$$

3.1.2 Extrapolation – Least Squares

By applying the Least Squares extrapolation method the assumption is made that PDs are cumulative Weibull distributed with an integrated ceiling parameter, as described in Equation 3-2. The ceiling parameter is implemented because the PD value does not approximate 1 when estimation period approximates infinity, due to the competing risk of prepayment, as explained in paragraph 2.3.

$$PD_{it} = 1 - c_i * e^{-\left(\frac{t}{\tau_{group}}\right)^{\alpha_i}}, \quad with\ i = customer, t = period\ and\ \ c = ceiling\ parameter \tag{3-2}$$

The first step of the Least Squares extrapolation method is to determine the parameters of the distribution $(c, \alpha, \tau_{group})$ for the different customer groups $(Healthy, Recovered, Arrears)$, with the help of the Least Squares regression technique (Bretscher, 1995). The Least Squares regression technique is a proven technique in solving systems with more equations that unknowns. A graphical representation of this first step is given in Figure 3.2, for a group of 15 Healthy customers.

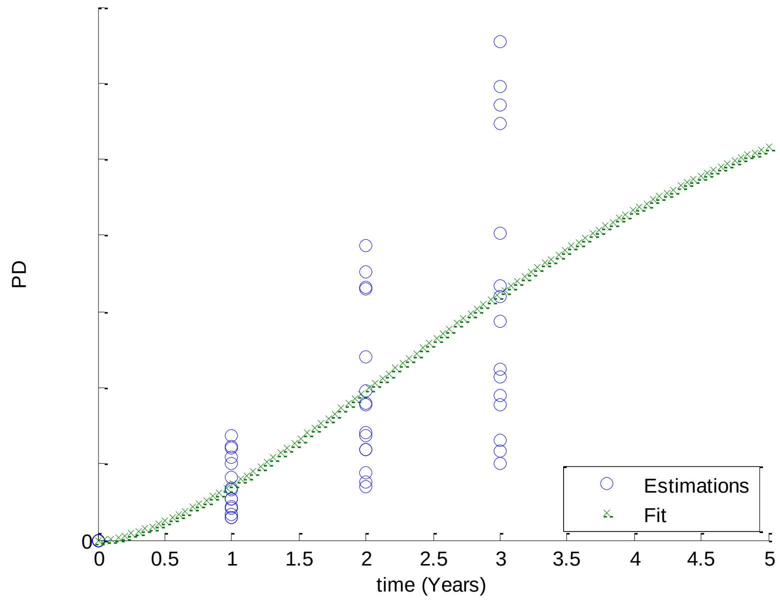


Figure 3-2 Example: Least Squares fit on 15 Healthy customers

The second step is to fix the scale parameter of the cumulative default distribution (τ) and determine the other parameters on a customer level. The scale parameter is fixed, because it encloses important group characteristics and it is assumed that the scale of the cumulative PD distribution is the same for customers that belong to the same group. A graphical representation of the second step is given in Figure 3.3, for a group of 15 Healthy customers. It can be seen in the figure that the cumulative default probabilities of the unique customers, found through estimation, matches with the direct 5 year Logit PD estimations per customer. With the parameters that results from the both steps, the LT PD will be calculated per customer.

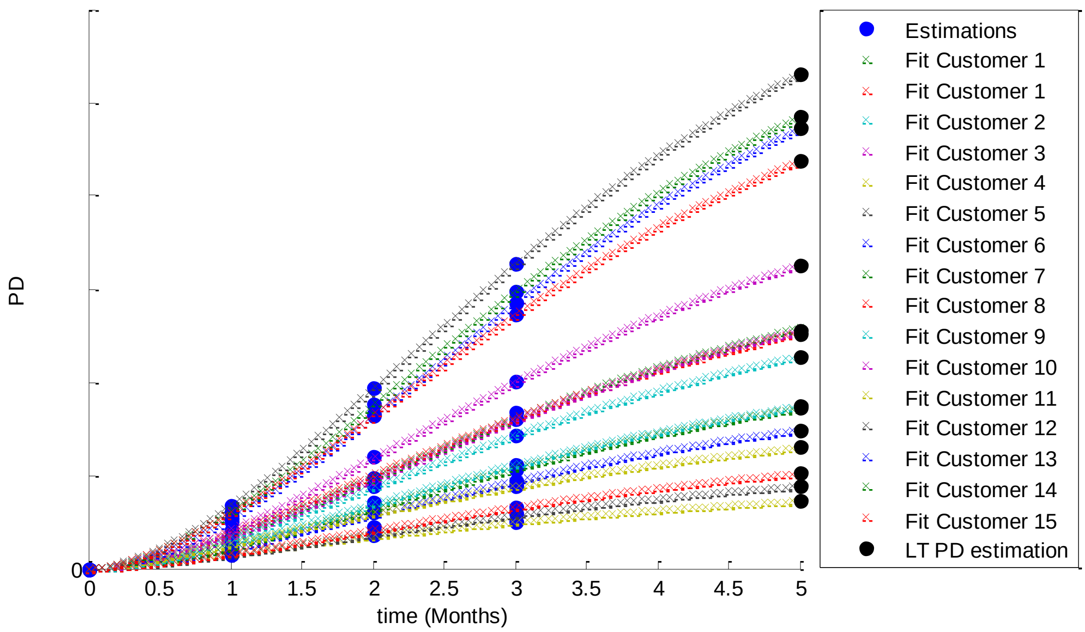


Figure 3-3 Example: Individual Least Squares fit for 15 Healthy customers

3.1.3 Extrapolation - PD given 1 year PD

The third method that will be evaluated is calculating the LT PD given the 1 year PD. The formula used for calculating a 5 year PD given a 1 year PD is illustrated with Equation 5-1. This method assumes that the 1 year probability of default and the 1 year probability of prepayment are representative for the latter years. If this is not the case, under and over estimation will occur.

$$5\ year\ PD = 1 - (1 - 1\ year\ PD)^{5/1} \tag{5-1}$$

3.2 Internal policy

The Internal Policy requirement is a requirement that follows from Basel regulation and is formulated as follows: *The criteria for the classification of loans have to fit with the internal policy and the policy of the department "Arrears management".* To meet this requirement a distinction is made between different customers. All customers are, based on selected characteristics, divided into three groups, named: Healthy, Recovered and Arrears, based on some of their characteristics. These characteristics are presentenced per group in Table 3-2. The three groups represent respectively around 92%, 6% and 2% of the total amount of customers.

3-2 Customer Groups and their Characteristics

Group	Characteristics
Healthy	# of arrears in the last 12 months = 0 and # of defaults in the last 12 months = 0
Recovered	Customer is not in arrears and # of arrears in the last 12 months > 0 or # of defaults in the last 12 months > 0
Arrears	Customer is in arrears

The distinction made is the same distinction as used in the PHIRM 2.0 PD model. During the development of the PHIRM 2.0 PD model an expert session took place in which the experts concluded that the distinction of customers in the three groups fits well with the internal policy and processes (SNS Bank N.V., 2013). Especially because the department "Arrears management" uses the same customer characteristics to distinct customer groups.

The result of the distinction in subgroups is that each prototype will consist of three 'sub' prototypes which are developed with the focus on estimating PDs for a specific customer group. Because this focus is in place, the sub models can have different variables.

By realising the prototypes according to the requirements as set by the internal policy and the policy of the department "Arrears management", also requirement 11[5] will be met.

[5] APPENDIX 3: *The model has to represent the current situation, not the situation how it should be. Inefficient processes also have to be modelled.*

4 Prototypes

This Chapter contains a description of the prototypes and sub prototypes based on the frameworks described in chapter 3. The process of variable selection (from the dataset described in chapter 5, which contains 137 variables) is described for each prototype. Also the selected variables and the estimated weights are presented. Because monthly observations on customer level are used, all prototypes can be calculated on a monthly and customer basis, whereby requirement 10 is fulfilled[6].

4.1 Logit Prototype

4.1.1 Variable selection

The variables of the Logit Prototype are selected with the help of the Credit Risk Interactive Modelling Environment (CRIME) tool, which is available inside the SNS Bank N.V. The CRIME tool alternately estimates a sub prototype with another variable and ranks the variables on added value by minimization of the Akaike Information Criterion (AIC) (Akaike, 1974) and calculates the weights and constant factor. When the user adds a variable, the CRIME tool estimates again all different variables for the sub prototype with the previously added variable. The CRIME tool prevents the selection of strong correlated variables, because after selecting one of two strong correlated variables the other will be at the bottom of the list next round. The CRIME tool divides the development dataset in 2 parts for purposes of estimation and validation to prevent over fitting[7].

The selection of the variables is done based on a heuristic that is developed in consultation with the external supervisors, because of their proven experience. The Logit prototypes, are as result of using a heuristic, statistical models. Always one of the three best variables regarding AIC have to be added, while trying to get an equal spread regarding the loan, pledge and customer characteristics. The process of variable selection stopped at the moment that the added value as result of adding an extra variable, expressed as AUROC, does not improve more than 0.5%, because this does not outweigh the added complexity of the prototype. Another rule followed is that the model must consist out of at least 3 variables and at most 6 variables, this to put in enough empirical evidence and to avoid complexity. This heuristic makes that the Logit Prototype meets requirements 1, 3[8] and 7[9]. The heuristic used is illustrated in a flowchart in Figure 4.1.

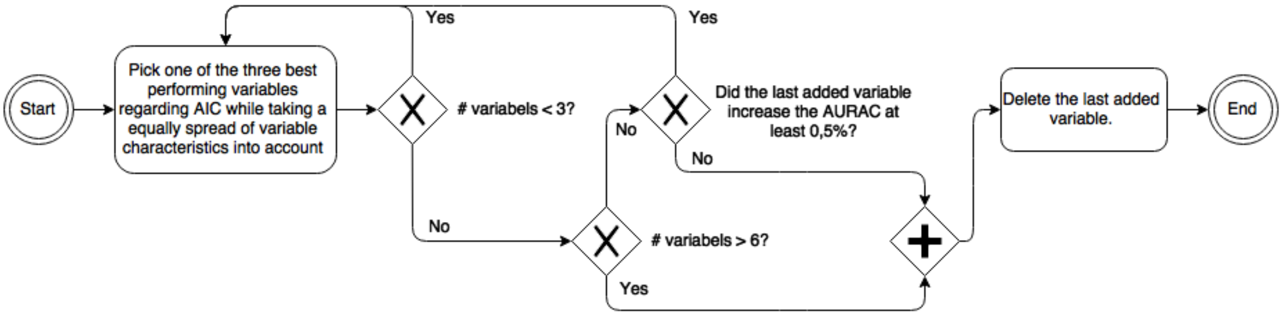


Figure 4-1 Flowchart of variable picking process

[6] APPENDIX 3: *The PD estimations have to be calculated at customer level and the model should enable the user to calculate them on a monthly basis.*
[7] More on this topic can be found in Chapter 5.
[8] APPENDIX 2: *The model has to contain borrower-, loan- and pledge characteristics, under the condition that that they have sufficient predictive power.*
APPENDIX 2: *Estimations of the PD have to be based on experience and empirical evidence and not on purely subjective considerations. The estimations also have to be intuitive and based on all relevant information. The less information is incorporated, the more conservative the estimates should be.*
[9] APPENDIX 3: *Used variables have to be selected based on their added value.*

4.1.2 Variables and their estimated weights

All selected variables for the sub prototypes of the 1 year Logit prototype, which will be used in Equation 2-1 to calculate PDs, are listed in Table 4-1. In this table the number of observations used in the train set[10], the train/test set ratio, the information about scaling, the characteristic type, the variable description and the value of the estimated weights are given. The variables selected for the $2, 3\ and\ 5\ year$ models and the variable names used inside SNS Bank N.V. can be found in APPENDIX 6.

Table 4-1 Logit Variables 1 year prototype

Sub Prototype, # of observations, Train/Test ratio, Scaling information	Characteristic type	Variable description	Estimated Value
Healthy, n = 581388, (70/30), X variable normalized	Borrower	Total number of arrears (#)	$\hat{\beta}_1$ = +0.20157
		Number of defaults in the last 36 months (#)	$\hat{\beta}_2$ = +0.0936576
	Loan	Average interest (%)	$\hat{\beta}_3$ = +0.27011
	Pledge	Consumer Credit part of extra mortgage (€)	$\hat{\beta}_4$ = -0.122986
		NHG principal (€)	$\hat{\beta}_5$ = -0.217029
		Loan to foreclosure value (%)	$\hat{\beta}_6$ = +0.499391
	Constant	-	$\hat{\beta}_0$ = -5.7205
Recovered, n=32422, (70/30), X variable normalized	Borrower	Number of arrears in the last 6 months (#)	$\hat{\beta}_1$ = +0.334063
		Total number of arrears (#)	$\hat{\beta}_2$ = +0.223352
		Number of defaults in the last 36 months (#)	$\hat{\beta}_3$ = +0.327995
	Loan	Average interest percentage (%)	$\hat{\beta}_4$ = +0.277961
	Constant	-	$\hat{\beta}_0$ = -2.6982
Arrears, n= 13184, (70/30), X variable normalized	Borrower	# defaults in the last 36 months (#)	$\hat{\beta}_1$ = +0.385075
		Amount arrears / month term (€)	$\hat{\beta}_2$ = -0.689566
	Loan	Payment arrangement indicator {0,1}	$\hat{\beta}_3$ = -0.319641
	Pledge	Loan to foreclosure value (%)	$\hat{\beta}_4$ = +0.153832
	Constant	-	$\hat{\beta}_0$ = -1.0931

4.2 Tobit Prototype

4.2.1 Variable selection

The variables of the Tobit Prototype are also selected with the help of the CRIME tool. The process of selection is almost the same as the process used to select the Logit variables. The only difference is the decision to stop the process which in this case is based on an adjusted version of the AUROC. Based on the process of variable selection it can be concluded that the Tobit Prototype meets requirements 1, 3[11] and 7[12].

[10] More information and explanation about the train and test set is given in chapter 5.
[11] APPENDIX 2: *The model has to contain borrower-, loan- and pledge characteristics, under the condition that that they have sufficient predictive power.*
APPENDIX 2: *Estimations of the PD have to be based on experience and empirical evidence and not on purely subjective considerations. The estimations also have to be intuitive and based on all relevant information. The less information is incorporated, the more conservative the estimates should be.*
[12] APPENDIX 3: *Used variables have to be selected based on their added value.*

4.2.2 Variables and their estimated weights

All selected variables for the sub prototypes of the 1 year Tobit prototype, which will be used in Equation 2-4 to calculate PDs, are listed in Table 4-2. In this table the number of observations used in the train set[13], the train/test set ratio, the information about scaling, the characteristic type, the variable description and the value of the estimated weights are given. Unfortunately the theoretical idea and the assumptions made to use the Tobit framework for estimating PDs, did during the research turn out to fail in practice. Therefore no multiyear Tobit models were created. Problems that occurred during the practical application of Tobit framework are analysed and provided with a direction of solution in chapter 6. The variable names used inside SNS Bank N.V. are presented in APPENDIX 7.

Table 4-2 Tobit Variables 1 year prototype

Sub Prototype, # of observations, Train/Test ratio, Scaling information	Characteristic type	Variable description	Estimated Value
Healthy, n = 396176, (50/50), X variable Scaled	Borrower	Total number of arrears (#)	$\hat{\beta}_1 = +0.711965$
	Loan	Term expired (months)	$\hat{\beta}_2 = -0.657999$
		NHG indicator {0,1}	$\hat{\beta}_3 = -0.282072$
	Pledge	Loan to foreclosure value (%)	$\hat{\beta}_4 = +0.331212$
	Constant	-	$\hat{\beta}_0 = -7.5003$
	Sigma	-	$\hat{\sigma} = 3.9756$
Recovered, n = 21898, (50/50), X variable Scaled	Borrower	Number of arrears in the last 6 months (#)	$\hat{\beta}_1 = +0.493453$
		Total number of arrears (#)	$\hat{\beta}_2 = +0.508115$
		Number of defaults in the last 36 months (#)	$\hat{\beta}_3 = +0.265358$
		Average amount arrears in the last 12 months (€)	$\hat{\beta}_4 = +0.130338$
	Loan	Arrears management loan indicator {0,1}	$\hat{\beta}_5 = +0.131762$
	Pledge	Loan to foreclosure value (%)	$\hat{\beta}_6 = +0.164622$
	Constant	-	$\hat{\beta}_0 = -0.22569$
	Sigma	-	$\hat{\sigma} = 2.4920$
Arrears, n = 8963, (50/50), X variable Scaled	Borrower	Number of defaults in the last 36 months (#)	$\hat{\beta}_1 = +0.506033$
		Amount arrears / month term (€)	$\hat{\beta}_2 = -0.764544$
	Loan	Arrears management loan indicator {0,1}	$\hat{\beta}_3 = +0.302402$
	Pledge	Loan to foreclosure value (%)	$\hat{\beta}_4 = +0.130984$
	Constant	-	$\hat{\beta}_0 = 2.2212$
	Sigma	-	$\hat{\sigma} = 2.4124$

4.3 Hazard Prototype

4.3.1 Variable Selection

The CRIME tool used for the selection of variables for the Logit- and Tobit frameworks unfortunately does not have the functionality to select variables for a Hazard prototype. Manually evaluating all possible combinations of the 137 variables takes too much time and therefore it has been decided to select the

[13] More information and explanation about the train and test set is given in chapter 5.

Hazard variables by the use of an heuristic. All variables that are used at least twice in the different Logit prototypes are selected. The Logit variables offer a stable basis, because they have proven to work in the Logit models. Selecting variables that are used at least twice in the Logit prototypes result in the selection of variables that have predictive and discriminatory power for multiple periods and/or customer groups. Furthermore a variable is added which represents the historic number of defaults. This variable is added to make a distinction in observations based on the place an observation has in the sequence of observations per customer, as result of the way the dataset is constructed, explained in chapter 5. The selected variables are used for the maximum likelihood optimization to calculate the weights.

Based on the process of variable selection used, it can be concluded that the Hazard Prototype meets requirements 1 and 3[14]. The hazard model does not meet requirement 7[15], which requires that the variables need to be selected on basis of added value, because the used approach does not give understanding in the added value of individual variables. When the hazard prototype seems to be the best or a promising prototype, an extensive new estimation process could be executed where after requirement 7 will be met.

4.3.2 Variables and their estimated weights

All selected variables for the sub prototypes of the Hazard prototype, which will be used in Equation 2-14 to calculate PDs, are listed in Table 4-3. In this table the Sub Prototype, the number of observations, the scaling information, the event, the characteristic type, the variable name and the value of the estimated weights are given. Because all the variables are selected by the use of a heuristic, all observations are used in the train set[16] for estimating the weights. The variable names used inside SNS Bank N.V. are presented in APPENDIX 8.

[14] APPENDIX 2: *The model has to contain borrower-, loan- and pledge characteristics, under the condition that that they have sufficient predictive power.*
APPENDIX 2: *Estimations of the PD have to be based on experience and empirical evidence and not on purely subjective considerations. The estimations also have to be intuitive and based on all relevant information. The less information is incorporated, the more conservative the estimates should be.*
[15] APPENDIX 3: *Used variables have to be selected based on their added value.*
[16] More information and explanation about the train and test set is given in chapter 5.

Table 4-3 Hazard Variables

Sub Prototype, # of observations, Scaling information	Event	Character-istic type	Variable description	Estimated Value
Healthy, n = 226290, X variables scaled on [0,1]	Default	Borrower	Total number of defaults (#)	$\hat{\beta}_1$ = -11.551969
			Total number of arrears (#)	$\hat{\beta}_2$ = -4.220677
			Number of defaults in the last 36 months (#)	$\hat{\beta}_3$ = 1.777056
		Loan	Average interest (%)	$\hat{\beta}_4$ = -1.193231
		Pledge	Consumer Credit part of extra mortgage (€)	$\hat{\beta}_5$ = -0.070829
			NHG principal (€)	$\hat{\beta}_6$ = 0.586302
			Loan to foreclosure value (%)	$\hat{\beta}_7$ = -3.342875
		Constant	-	$\hat{\beta}_0$ = 9.217739
		Alpha	-	$\hat{\alpha}_{def}$ = 0.761171
	Prepay-ment	Tau	-	$\hat{\tau}_{prep}$ = 209.641623
		Alpha	-	$\hat{\alpha}_{prep}$ = 1.340222
Recovered, n = 17817, X variables scaled on [0,1]	Default	Borrower	Total number of defaults (#)	$\hat{\beta}_1$ = -6.645824
			achter_3mnd_ind	$\hat{\beta}_2$ = -0.399239
			Total number of arrears (#)	$\hat{\beta}_3$ = -1.924554
			Number of defaults in the last 36 months (#)	$\hat{\beta}_4$ = -0.260921
			Number of arrears in the last 6 months (#)	$\hat{\beta}_5$ = -0.864793
		Loan	Average interest (%)	$\hat{\beta}_6$ = -1.74342
		Pledge	Loan to foreclosure value (%)	$\hat{\beta}_7$ = -1.139098
		Constant	-	$\hat{\beta}_0$ = 7.245091
		Alpha	-	$\hat{\alpha}_{def}$ = 0.835663
	Prepay-ment	Tau	-	$\hat{\tau}_{prep}$ = 221.297099
		Alpha	-	α_{prep} = 1.097646
Arrears, n = 5587, X variables scaled on [0,1]	Default	Borrower	Total number of defaults (#)	$\hat{\beta}_1$ = -5.283660
			Number of defaults in the last 36 months (#)	$\hat{\beta}_2$ = -0.058718
			Amount arrears / month term (€)	$\hat{\beta}_3$ = 1.291860
			Number of arrears in the last 12 months (#)	$\hat{\beta}_4$ = -0.945830
		Loan	Arrears management loan indicator {0,1}	$\hat{\beta}_5$ = -0.524896
			Payment arrangement indicator {0,1}	$\hat{\beta}_6$ = 0.1866110
		Pledge	Loan to foreclosure value (%)	$\hat{\beta}_7$ = -1.253996
		Constant	-	$\hat{\beta}_0$ = 4.406391
		Alpha	-	$\hat{\alpha}_{def}$ = 0.708987
	Prepay-ment	Tau	-	$\hat{\tau}_{prep}$ = 208.118199
		Alpha	-	$\hat{\alpha}_{prep}$ = 1.2442940

5 Description of the Data

This chapter contains a clear description of the data used and in what way the data is used.

5.1 Dataset

The dataset used during this Thesis is an edited version of two other datasets and contains monthly records of all for this Thesis relevant 326298 unique retail mortgage customers, who had a mortgage between September 2007 and September 2014. The data set contains the same variables that are used during the development of the PHIRM PD 2.0 model and therefore all frameworks will meet the 8th requirement[17]. The outliers in the dataset were already detected and replaced by outlier boundaries before the start of this Thesis. A description of this process is available for insiders at the SNS Bank N.V (A. Holst, 2014). Furthermore the DBV portfolio is filtered as well as the fraudulent/incorrect observations. Except the DBV portfolio is the dataset representative for all other portfolio's, and therefore will all frameworks meet requirement 9[18]. More specific information about the dataset, only interesting for insiders at the SNS Bank N.V. can be found in APPENDIX 4.

5.2 Subsets

For purposes of development and testing two different data sets are created. The development set, or train set, contains 70% of the customer related observations from the dataset described above and will be used to develop the prototypes. Those observations are randomly assigned. The test set, or Out of Sample set (OOS), contains the other 30% of the customer related observations and will be used for testing purposes. By dividing the dataset in two subsets the user of CRIME can detect over fitting.

Other separations are made regarding the period of observation. The 1 year prototypes are developed by using data from the period September 2007 till September 2012. The LT prototypes are developed by using data from the period September 2007 till September 2008. The reason to do this is that a forward looking indicator is needed for estimating purposes and a backward looking indicator is used for dividing the observations in sections. By making the interval smaller (of the total available reliable data September 2006 till September 2014) 'forward and backward data' becomes available. The period following the development period will be used as Out of Time set (OOT), to assess the performance of the prototypes on a period of time that was not included during the development.

5.3 Prototype related data matters

5.3.1 Logit- and Tobit prototype

A sampling technique is used to tackle the seasonal effects in the data, before using the data to develop the Logit- and Tobit prototypes. Each unique customer in the dataset got a randomly assigned month of observation $(1,2,..,12)$. For each customer only observations are used that correspond with their month of observation.

Besides the sampling technique also the customers that were already in default are filtered. This because the PD only has to be defined for customers that are not in default yet. The PD for a customer that is already in default is equal to 1.

---

[17] APPENDIX 3: *The variables that may be used have to be part of the short list or have to be used in the current PHIRM 2.0 PD model.*

[18] APPENDIX 3: *The model may be fitted on a sub portfolio, but the performance has to be comparable for the complete portfolio. The DBV portfolio is out of scope in this project, due to the fact that it is not comparable with the other portfolios.*

5.3.2 Hazard prototype

To gather the data for the hazard prototype another data selection technique is needed. The hazard prototype needs for development and testing purposes combinations of risk indicators, an event (default or total prepayment), the corresponding censoring indicators, the age at the start of the observation period, the time in between the start of the loan and the start of the observation period and the time between the start of the observation period and the event. These combinations are made as illustrated for default events in Figure 5-1. The variables at the moment of time at the beginning of each combination $(3A, 3B \text{ and } 3C)$ are combined with the realization at the end of the combination $(Default,\ Total\ Preypayment\ or\ Censored)$. The first combination $(3A)$ does start earlier than the beginning of the observation period, so this observation will be corrected for age during estimation.

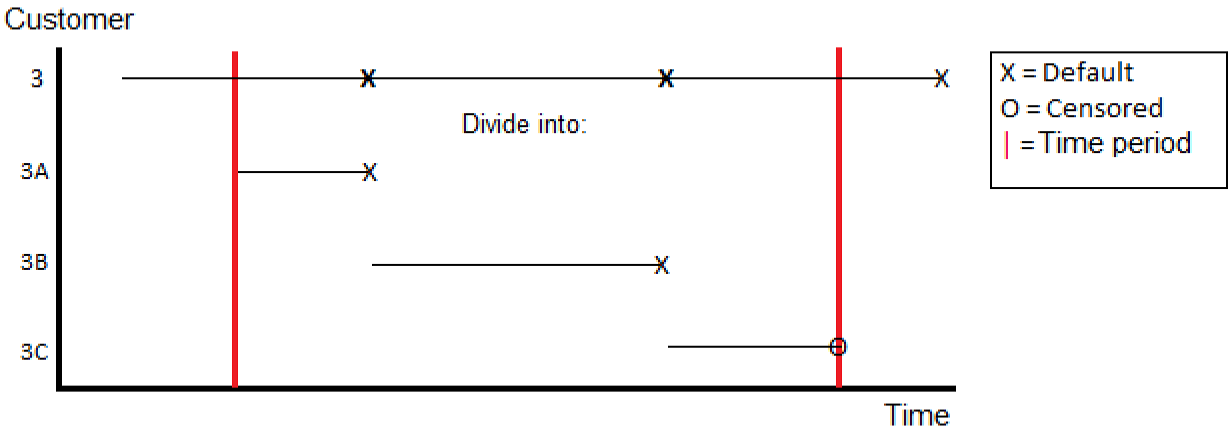


Figure 5-1 Hazard data selection technique

To estimate the hazard prototype also censored indicators are constructed for the events of Default and Total Prepayment. When an combination ends with a default event, the combination will censored for prepayment, because the information regarding prepayment of this combination is that the loan is not pre-paid till the event of default. When an combination ends with a prepayment event, the combination will censored for default, because the information regarding default of this combination is that the loan is not in default till the event of prepayment. When an combination runs out of the observation period, as is the case for combination $(3C)$, the combination will be censored for both default and prepayment, because the information regarding default and prepayment of this combination is that the loan is not in default and also not pre-paid till the event of prepayment. An example of the interaction of indicators and variables is given for three customers in the last table of APPENDIX 4.

Important to notice is that for this way of constructing the dataset, it is assumed that defaults could be reset, in the sense that after a default a new observation could start.

6 Results and Analysis

Chapter 6 contains the results of estimating PDs with the 3 different prototypes and an analysis is presented based on the methods according the 5th sub-question, presented in paragraph 2.4. For frameworks of which the practical application did not turn out as theoretically intended, the problems that occurred are analysed and provided with possible solutions.

6.1 Logit framework - Quantitative analysis & Results

Because there are in total three methods to estimate LT PDs with a Logit prototype, as described in paragraph 3.1, this quantitative analysis & results paragraph consists out of five sub-paragraphs. The first sub-paragraph gives insight in the performance of the 1 year Logit prototype. The following sub-paragraphs gives insight in the performance of the different methods to estimate LT PDs. The best performing LT method, the Least Squares method, will be used in the Logit LT prototype, which will be compared with the prototypes based on the other frameworks. All tests are performed on an OOS dataset, covering both OOT as non OOT observations. This paragraph ends by summarizing the results regarding the Logit 1 year and LT prototypes.

6.1.1 1 year Logit prototype

To gather meaningful insight in the predictive power of the 1 year Logit prototype, the estimated PDs by the model are plotted against the realized DR on a portfolio level in Figure 6.1. Important to notice is the vertical red line, which is the boundary of OOT and non OOT observations. From the figure can be concluded that the 1 year Logit prototype estimations follow the realized DR. Besides that, no seasonal effects or constant bias can be observed in the estimations, whereby requirements 4 and 6[19] are met. The absence of seasonal effects is the result of the used sampling method, explained in chapter 5.

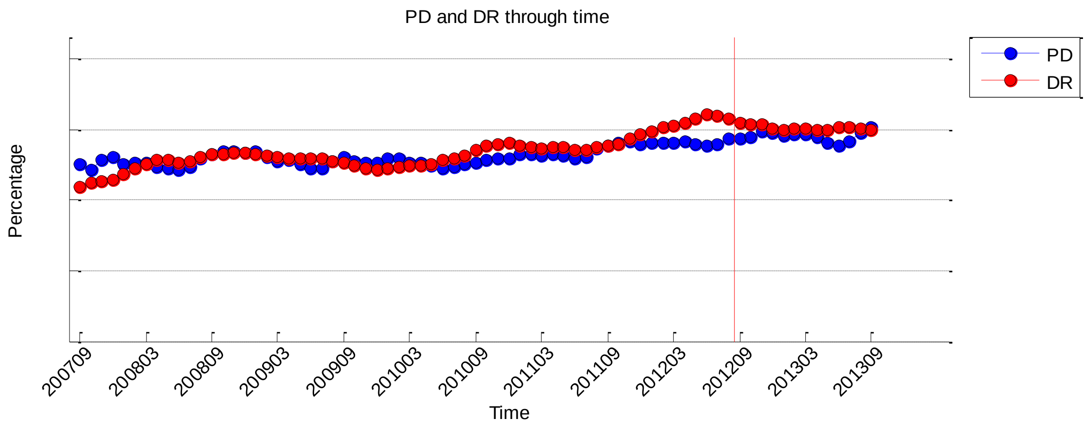


Figure 6-1 PD and DR through time - 1 year Logit prototype

Figure 6.2 shows that the Logit 1 year prototype also succeeds in having a good predictive power for different sections, whereby requirement 6 is met[20]. In the Figure the important sections $Healthy, Recovered\ and\ Arrears$ are shown. The other sections can be found in APPENDIX 9, together with visualized test results and visualized statistics, for example: the number of observations per section and moment of measurement.

[19] APPENDIX 2: Seasonal effects have to be observed and analysed.
APPENDIX 2: The model must have a good predictive power and no material deviation.
[20] APPENDIX 3: The model's predictive power has to be sufficient for different sections. These sections are: NHG/non NHG customers, Healthy/Recovered/Default and different geographical regions (East and West).

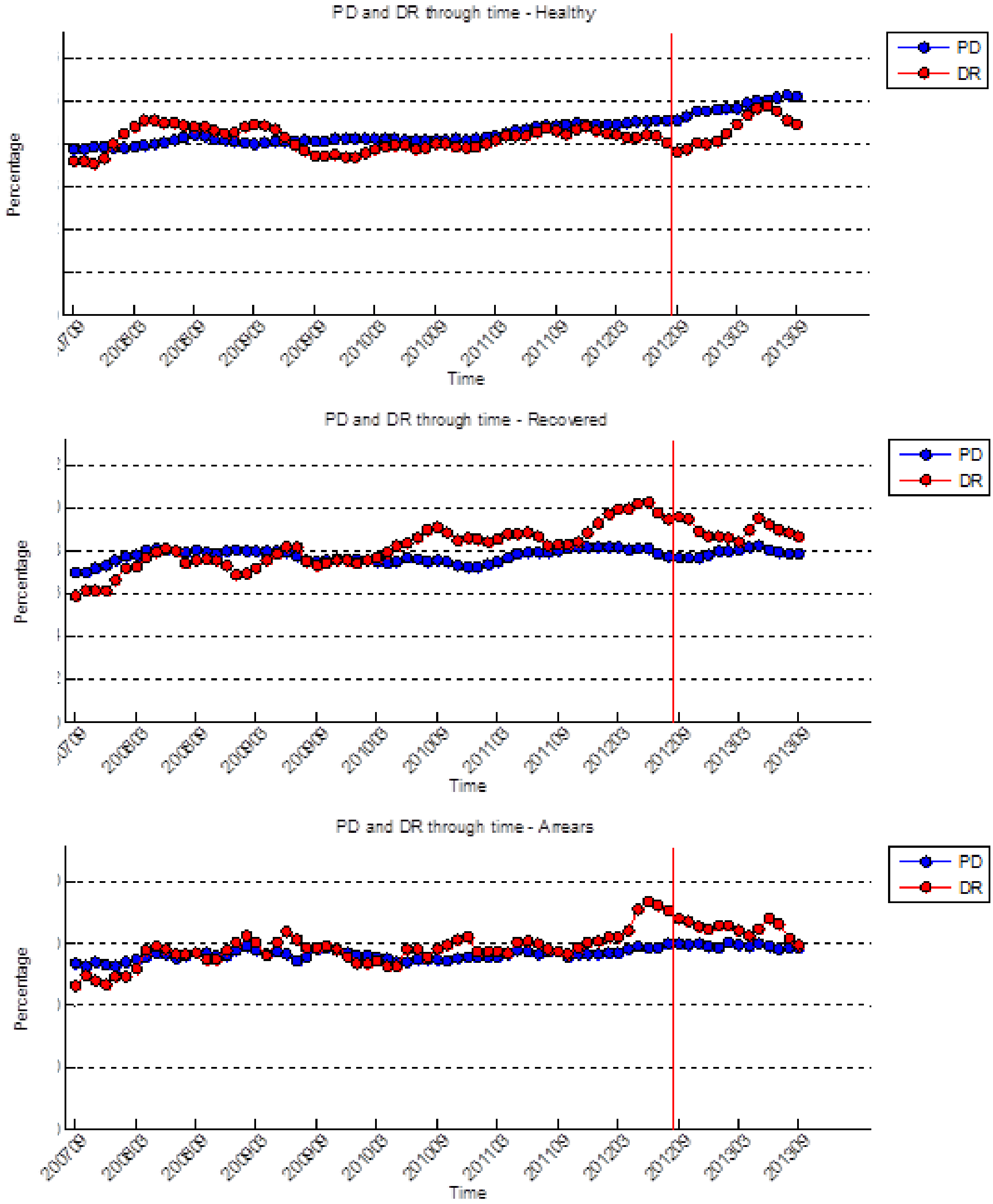


Figure 6-2 PD and DR through time for sections 'Healthy', 'Recovered' and 'Arrears' - 1 year Logit prototype

Figure 6.3 shows the results of the two sided binomial test on portfolio level with a significance of $\alpha = 0.05$. The black cubes indicate that the realization lies inside the confidence interval, a white cube indicates that this is not the case. When using the binomial test the assumption is made that the number of clients who will default follows a binomial distribution. Based on this assumption a confidence interval is created around the PDs. One should expect that in 95% of the cases, the realization would lie in this confidence interval. It is clearly that in this case the result does not match with the expectation. This because the Logit Prototype is not a PIT model, which results in failing the test as explained in paragraph 2.4.1. The results of the two sided binomial test on bucket level are presented in APPENDIX 9.

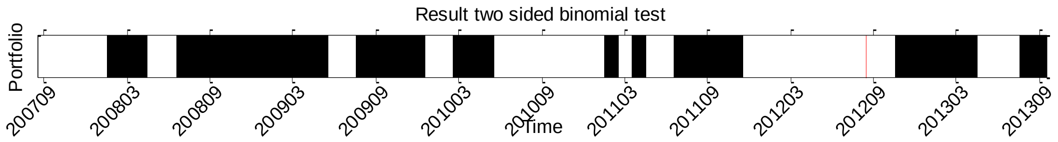


Figure 6-3 Result on portfolio level of the two sided binomial test- 1 year Logit prototype

The Normal Test hypotheses are rejected for bucket 4 till 10 (of the in total 12 buckets[21]). This could be the case because of an upward trend in DR, which results in OOT realizations which are all higher than the non OOT average DR. The Vasicek test passed for all buckets and on all moments of measurement.

To gather meaningful insight in the discriminatory power of the 1 year Logit prototype, the AUROC is plotted on portfolio level in Figure 6.4. From the figure can be concluded that the 1 year Logit prototype performs good till excellent regarding the discriminatory power[22]. Because of this requirement 5 is met[23].

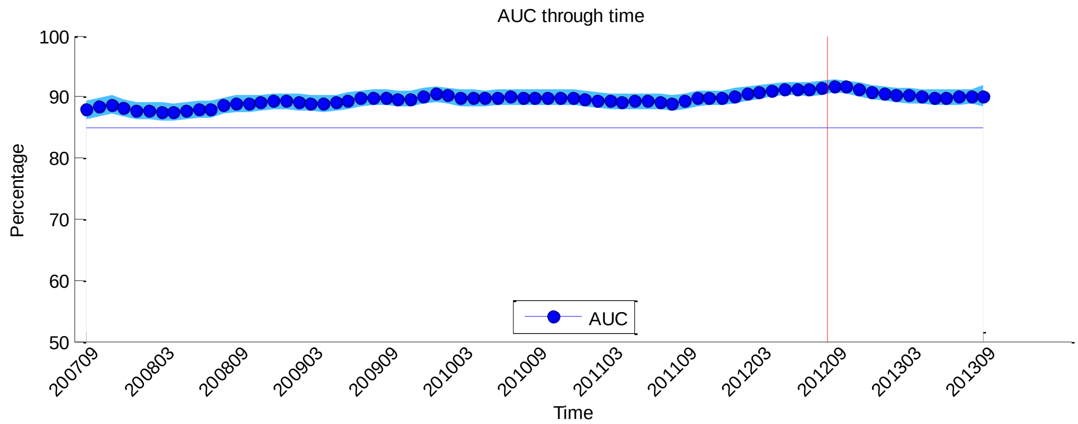


Figure 6-4 AUROC through time on portfolio level - 1 year Logit prototype

Figure 6.5 shows that it is harder for the Logit 1 year prototype to succeed in having a good discriminatory power for different sections. In the Figure the important sections $Healthy, Recovered\ and\ Arrears$ are shown and the results are moderate till sufficient and quite constant. The AUROC figures for the other sections are presented in APPENDIX 9.

[21] All buckets and bucket borders are presented in APPENDIX 5.
[22] All statements about discriminatory power are based on the scale introduced by Swets (Swets, 1988), which is presented in Chapter 2.
[23] APPENDIX 2: The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.

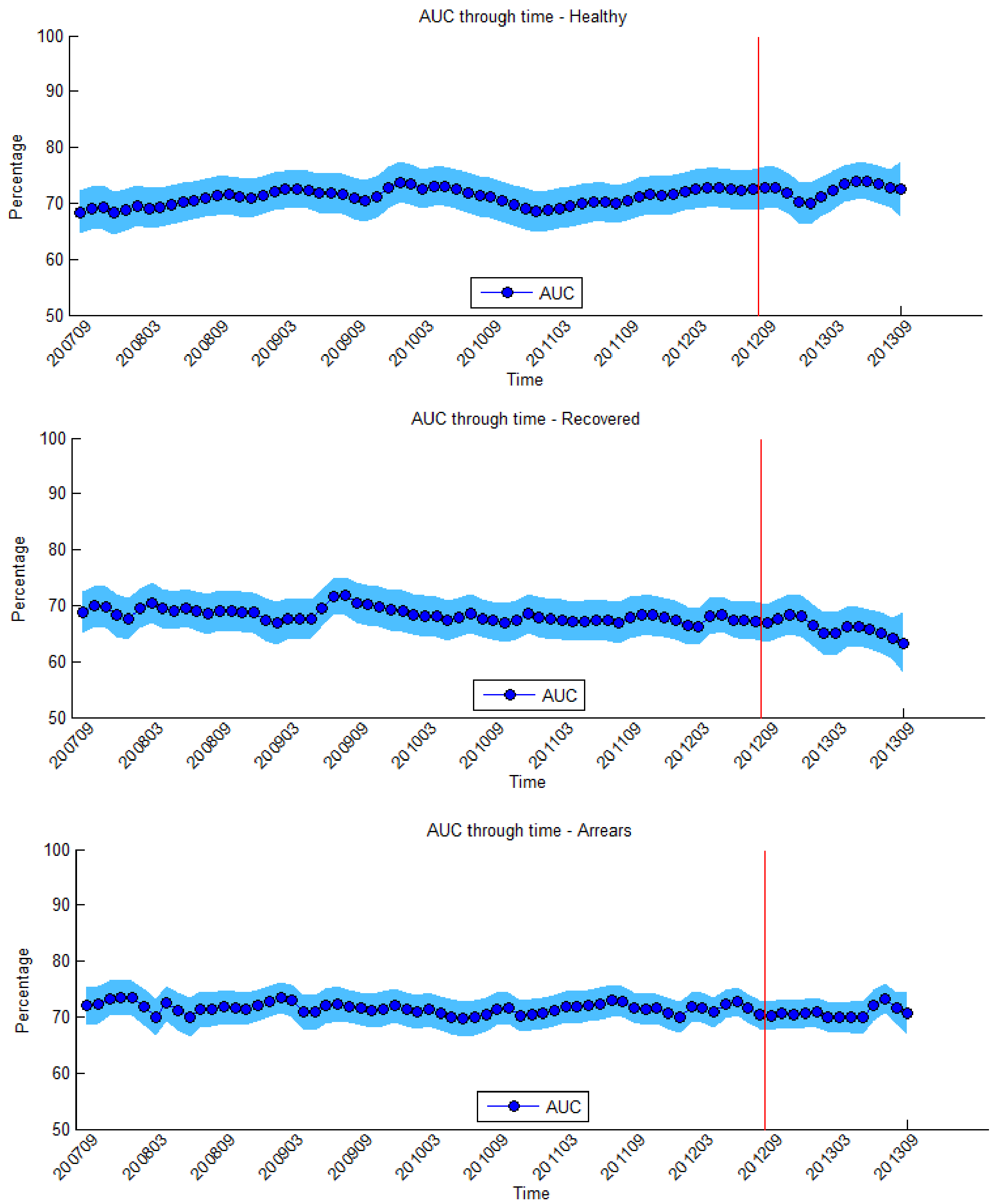


Figure 6-5 AUROC through time for sections 'Healthy', 'Recovered' and 'Arrears' - 1 year Logit prototype

The Kolmogorov-Smirnov test shows that PDs of defaulted customers estimated by the Logit 1 year prototype significantly differ from PDs estimated by the Logit 1 year prototype of not defaulted customers. The test hypothesis is accepted for every moment of measurement.

Concluding can be said that the 1 year Logit model meets all the quantitative requirements, it has a sufficient predictive power and also a sufficient discriminatory power.

6.1.2 Extrapolation – Mean Lambda

To gather meaningful insight in the predictive power of the LT Logit prototype with mean Lambda extrapolation, the estimated PDs are plotted against the realized DR on portfolio level in Figure 6.6. Important to notice is the vertical red line, which is the border of OOT and non OOT observations. From the figure can be concluded that the LT Logit prototype with mean Lambda extrapolation estimations follows the realized DR. Besides that, no seasonal effects or constant bias can be observed in the estimations, whereby requirements 4 and 6[24] are met on portfolio level. The absence of seasonal effects is the result of the used sampling method, explained in chapter 5.

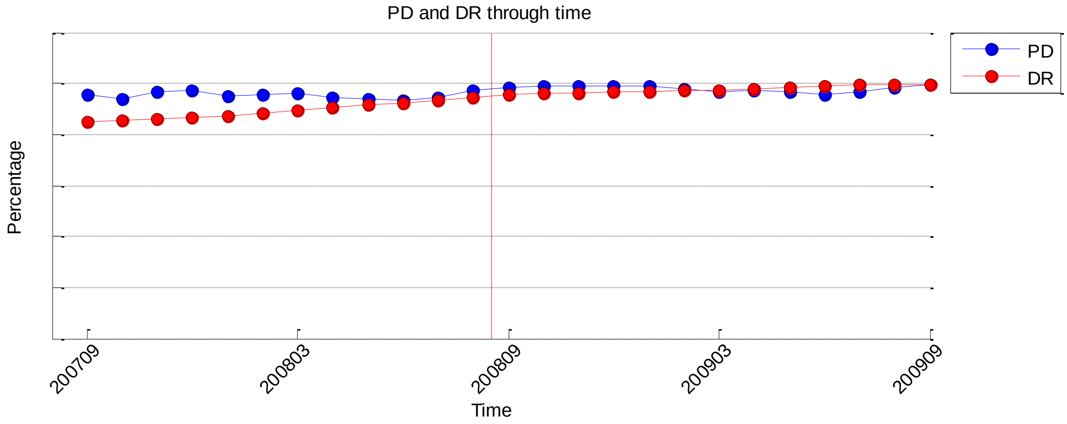


Figure 6-6 PD and DR through time - LT Logit prototype with mean Lambda extrapolation

Figure 6.7 shows that the LT Logit prototype with mean Lambda extrapolation fails in having a good predictive power on the different sections. A material bias, consisting of over- and under estimation, is observable for all the important sections: $Healthy, Recovered$ and Arrears. Therefore the LT Logit Prototype with mean Lambda extrapolation does not meet the requirements regarding predictive power on sectional level. The DR and PD plots of the other sections can be found in APPENDIX 10, together with visualized test results and visualized statistics, for example: the number of observations per section and moment of measurement.

[24] APPENDIX 2: Seasonal effects have to be observed and analysed.
APPENDIX 2: The model must have a good predictive power and no material deviation.

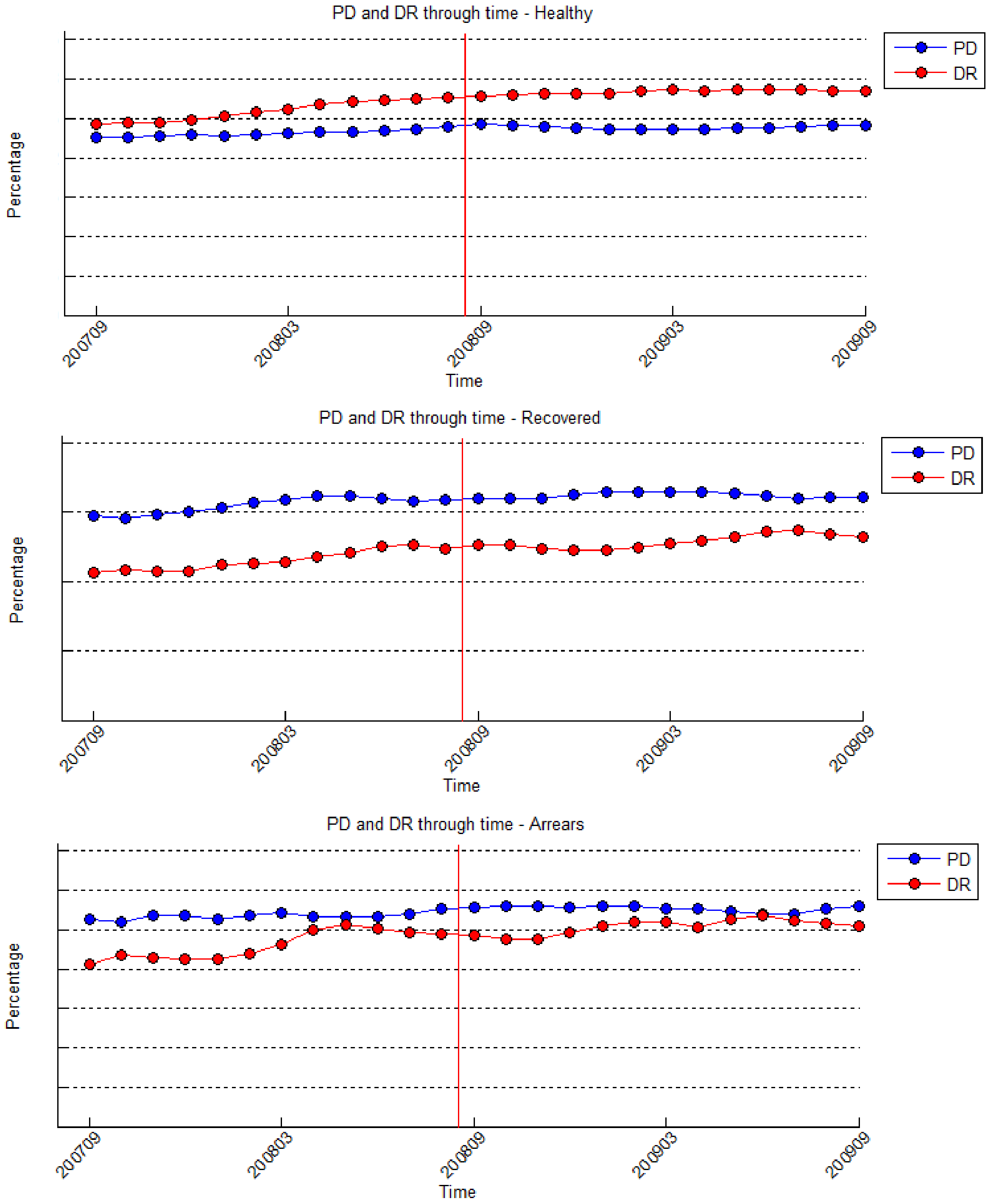


Figure 6-7 PD and DR through time for sections 'Healthy', 'Recovered' and 'Arrears' - LT Logit prototype with mean Lambda extrapolation

Figure 6.9 shows the results of the two sided binomial test on portfolio level with a significance of $\alpha = 0.05$. The black cubes indicate that the realization lies inside the confidence interval, a white cube indicates that this is not the case. It is clearly that the result does not match with the expectation that 95% of the realizations will lie in the confidence interval. This because the Logit prototype is not a purely TTC or PIT model, which results in failing the test as explained in paragraph 2.4.1. The results of the two sided binomial test on bucket level are presented in APPENDIX 10.

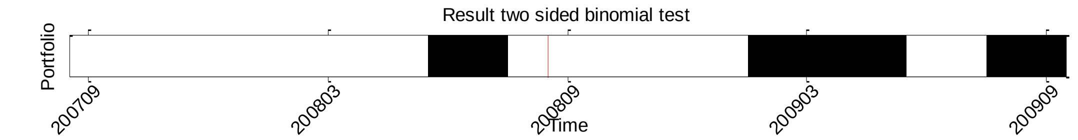


Figure 6-8 Result on portfolio level of the two sided binomial test- LT Logit prototype with mean Lambda extrapolation

The Normal Test hypotheses are rejected for bucket 7 till 9. This could be the case because of the upward trend in DR, which results in OOT realizations which are all higher than the non OOT average DR. The hypotheses of the Vasicek test are accepted for all buckets on all moments of measurement.

To gather meaningful insight in the discriminatory power of the LT Logit prototype with mean Lambda extrapolation, the AUROC is plotted a portfolio level in Figure 6.9. From the figure can be concluded that the Logit prototype with mean Lambda extrapolation performs good regarding the discriminatory power. Because of this requirement 5 is met on portfolio level[25].

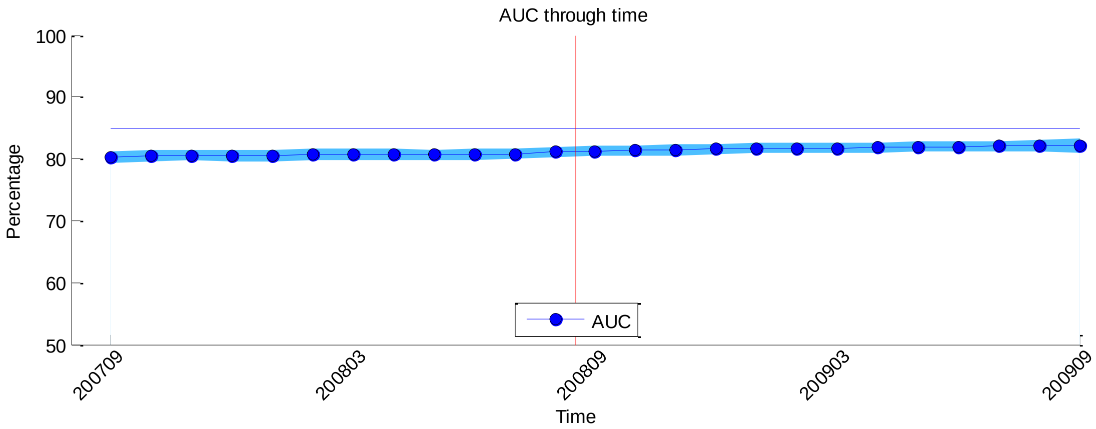


Figure 6-9 AUROC through time - LT Logit prototype with mean Lambda extrapolation

Figure 6.10 shows that it is harder for the LT Logit prototype with mean Lambda extrapolation to succeed in having a good discriminatory power for different sections. In the Figure the important sections $Healthy, Recovered\ and\ Arrears$ are shown and the results are weak till moderate and quite constant. Because of this requirement 5 is not met on sectional level[26]. The other AUROC for the other sections can be found in APPENDIX 10.

[25] APPENDIX 2: The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.
[26] See footnote 21.

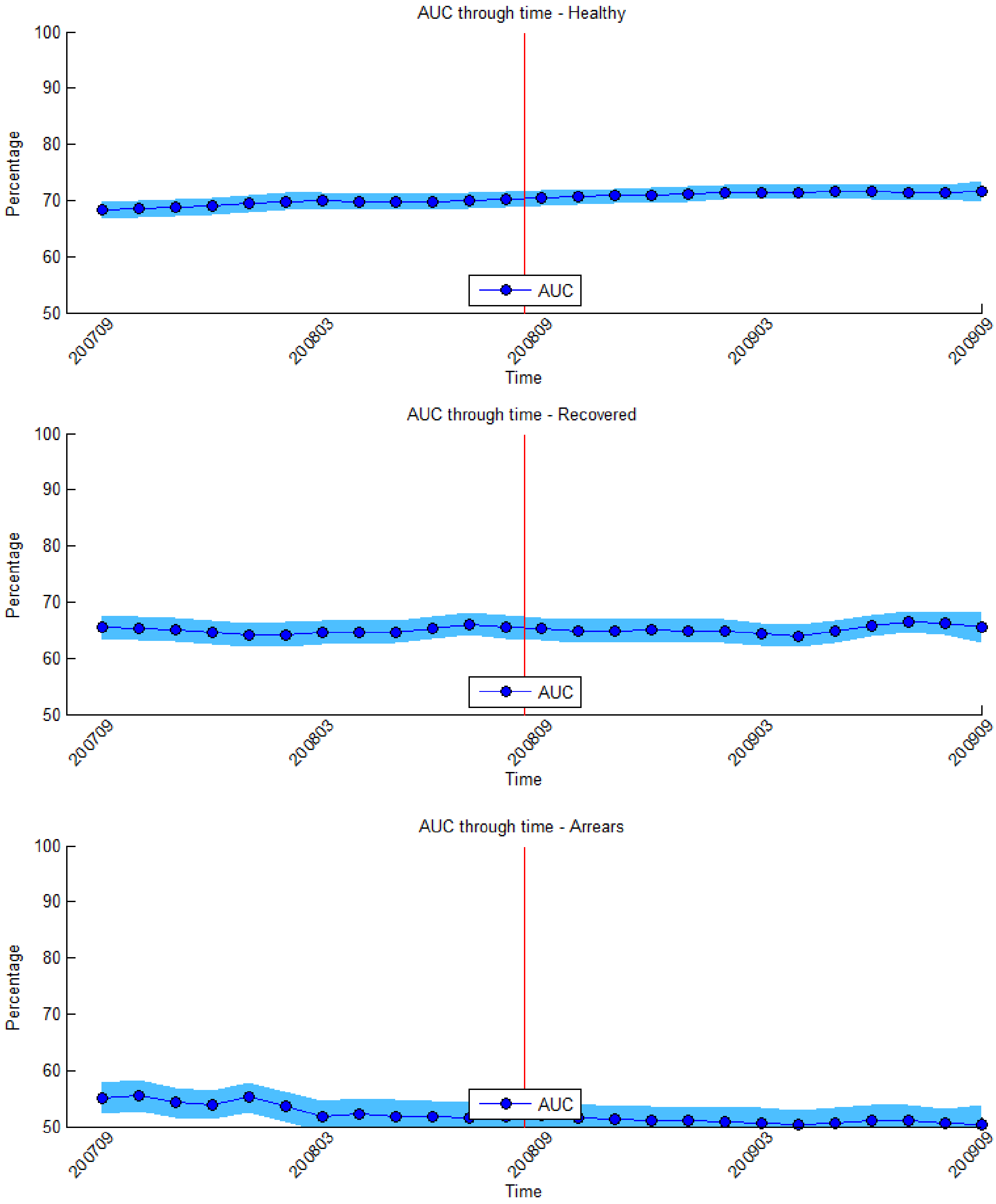


Figure 6-10 AUROC through time for sections 'Healthy', 'Recovered' and 'Arrears' - LT Logit prototype with mean Lambda extrapolation

The Kolmogorov-Smirnov test shows that PDs of defaulted customers estimated by the LT Logit prototype with mean Lambda extrapolation significantly differ from PDs estimated by the LT Logit prototype with mean Lambda extrapolation of not defaulted customers. The test hypothesis is accepted for every moment of measurement.

Concluding can be said that LT Logit prototype with mean Lambda extrapolation does not meet all the quantitative requirements. A material bias can be observed when the predictive power is visualized and dived in sections. Also the AUROC is much lower than the AUROC from the 1 year Logit prototype.

Experiments with a direct 5 year Logit prototype (without extrapolation) showed that it is not only the longer period that reduces the predictive and discriminatory power on the sections. The PD and DR of the direct 5 year prototype had a much smoother fit and an AUROC from moderate to sufficient on the different sections[27]. This indicates that the extrapolation method used does not work as intended. It could be the case that the assumption of the cumulative Exponential distribution is not correct. Another problem could be that the mean average Lambda of just three points does not represents the Lambda of the total distribution. In that case extra models could be made for different periods of time to get a better approximation of Lambda. Also a ceiling could be added to prevent the PD from going to 1 if the time goes to infinity. On a 5 year horizon the difference between the Least Squares method with, or without a ceiling, is small, on longer periods it is probable that this difference will be big.

6.1.3 Extrapolation – Least Squares

To gather meaningful insight in the predictive power of the LT Logit prototype with Least Squares extrapolation, the estimated PDs by the model are plotted against the realized DR on a portfolio level in Figure 6.11. Important to notice is the vertical red line, which is the border of OOT and non OOT observations. From the figure can be concluded that the LT Logit prototype with mean Lambda extrapolation estimations follows the realized DR. Besides that, no seasonal effects or constant bias can be observed in the estimations, whereby requirements 4 and 6[28] are met on portfolio level. The absence of seasonal effects is the result of the used sampling method, explained in chapter 5.

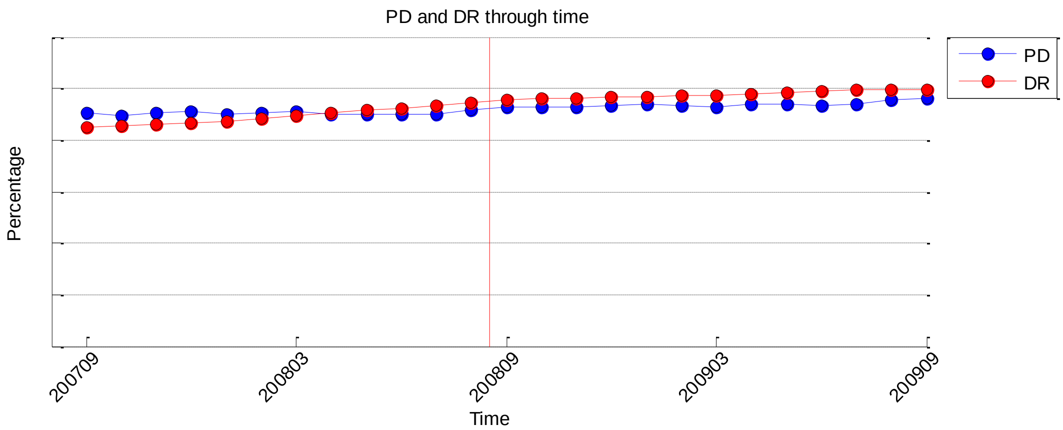


Figure 6-11 PD and DR through time - LT Logit prototype with Least Squares extrapolation

Figure 6.12 shows that the LT Logit prototype with Least Squares extrapolation fails in having a good predictive power on the sections, though the figures look better than from the LT Logit prototype with mean Lambda extrapolation. A material bias, consisting of over- and underestimation, is visible for all the important sections: *Healthy*, *Recovered* and Arrears. Therefore the LT Logit Prototype with Least Squares extrapolation does not met the requirements regarding predictive power on sectional level. The other sections, wherein the prototype performs better, can be found in APPENDIX 12, together with visualized test results and visualized statistics, for example: the number of observations per section and moment of measurement.

[27] See APPENDIX 11 for the Logit direct 5 year prototype PD versus DR and AUROC figures.
[28] APPENDIX 2: Seasonal effects have to be observed and analysed.
APPENDIX 2: The model must have a good predictive power and no material deviation.

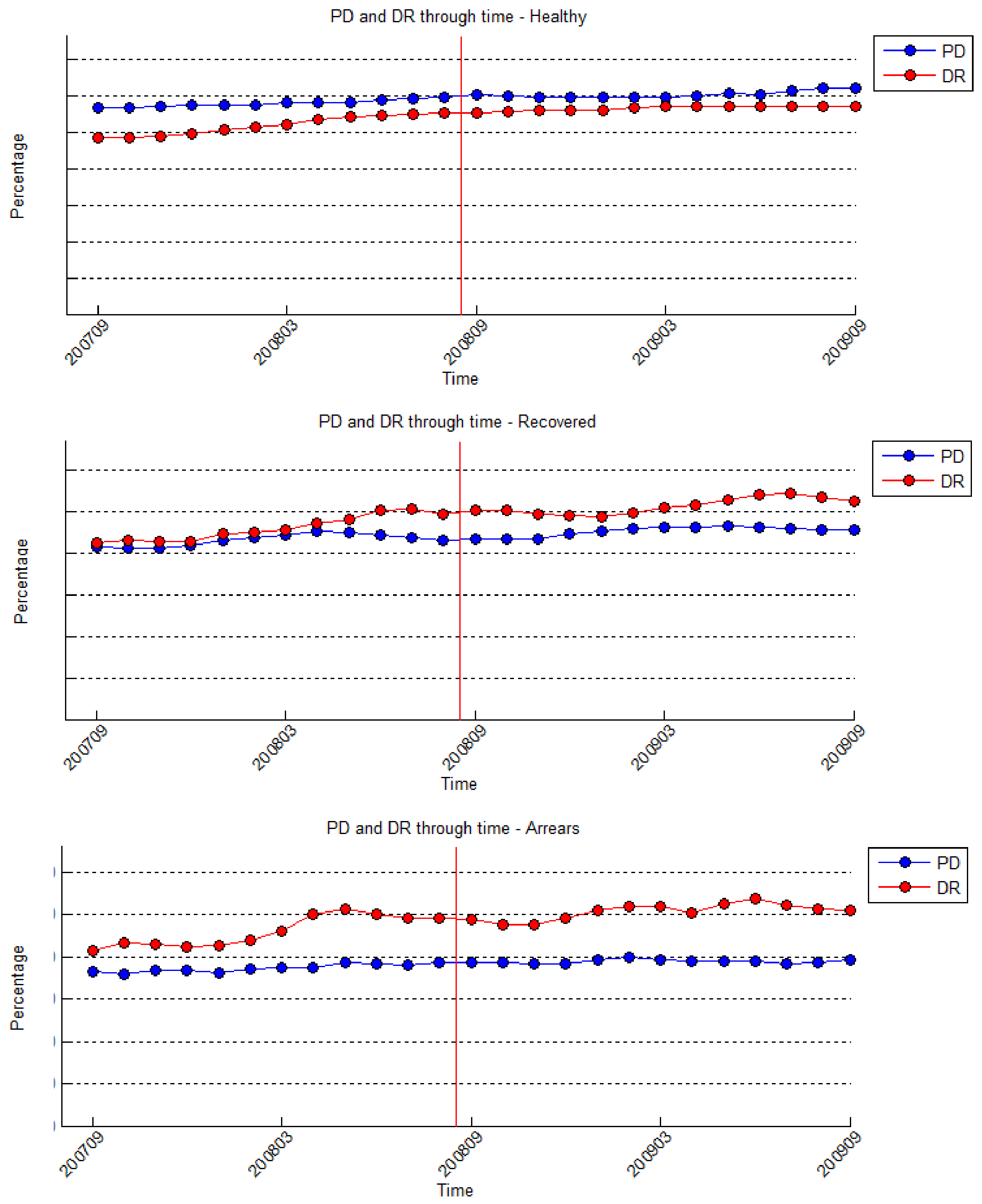


Figure 6-12 PD and DR through time for sections 'Healthy', 'Recovered' and 'Arrears' - LT Logit prototype with Least Squares extrapolation

Figure 6.14 shows the results of the two sided binomial test on portfolio level with a significance of $\alpha = 0.05$. The black cubes indicate that the realization lies inside the confidence interval, a white cube indicates that this is not the case. It is clear that the result does not match with the expectation that 95% of the realizations will lie in the confidence interval. This is probably the case because the Logit prototype is not a purely PIT model, which results in failing the test as explained in paragraph 4.2.1. Though the expectation is not met, the results of the two sided binomial test are better for the Logit prototype with Least Squares extrapolation than for the Logit prototype with mean Lambda extrapolation. The results of the two sided binomial test on bucket level are presented in APPENDIX 10.

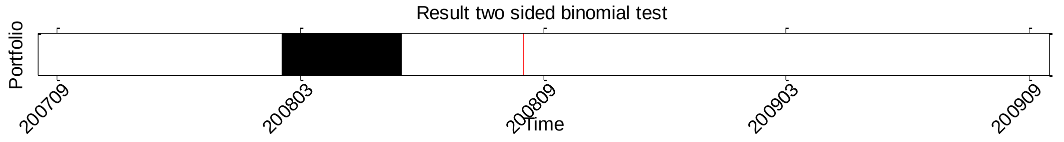


Figure 6-13 Result on portfolio level of the two sided binomial test- LT Logit prototype with Least Squares extrapolation

The Normal Test hypotheses are rejected for bucket 8 till 12. This could be the case because of the upward trend in DR, which results in OOT realizations which are all higher than the non OOT average DR. The hypotheses of the Vasicek test are accepted for all buckets on all moments of measurement.

To gather meaningful insight in the discriminatory power of the LT Logit prototype with Least Squares extrapolation, the AUROC is plotted a portfolio level in Figure 6.14. From the figure can be concluded that the Logit prototype with Least Squares extrapolation performs good regarding the discriminatory power. Because of this good performance, requirement 5 is met on portfolio level[29].

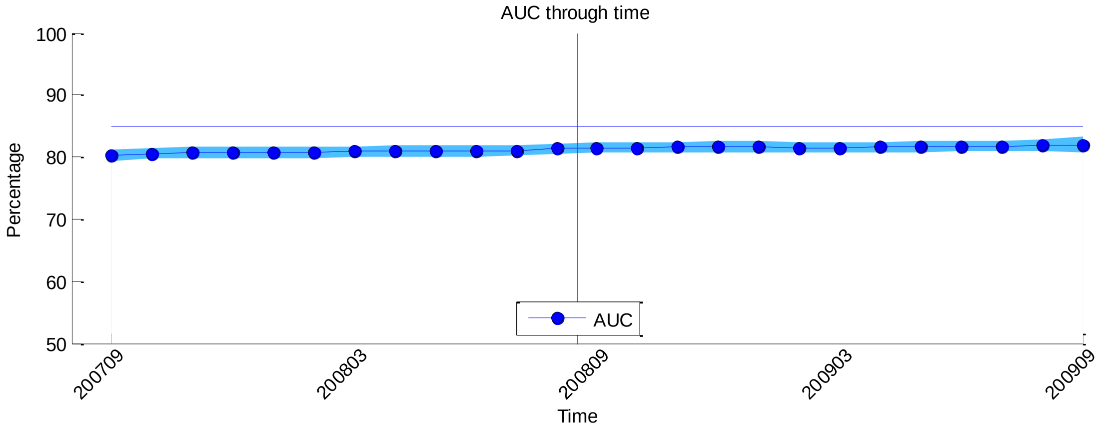


Figure 6-14 AUROC through time - LT Logit prototype with Least Squares extrapolation

Figure 6.15 shows that it is harder for the - LT Logit prototype with Least Squares extrapolation prototype to succeed in having a good discriminatory power on different sections. In the Figure the important sections: $Healthy, Recovered\ and\ Arrears$ are shown and the results are moderate till sufficient and quite constant. Because of this requirement 5 is not met on sectional level[30], but the results are better than for the LT Logit Prototype with mean Lambda extrapolation and also better than the results from experiments with the 5 year Logit prototype. The AUROC figures of the other sections can be found in APPENDIX 12.

[29] APPENDIX 2: The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.
[30] See footnote 24.

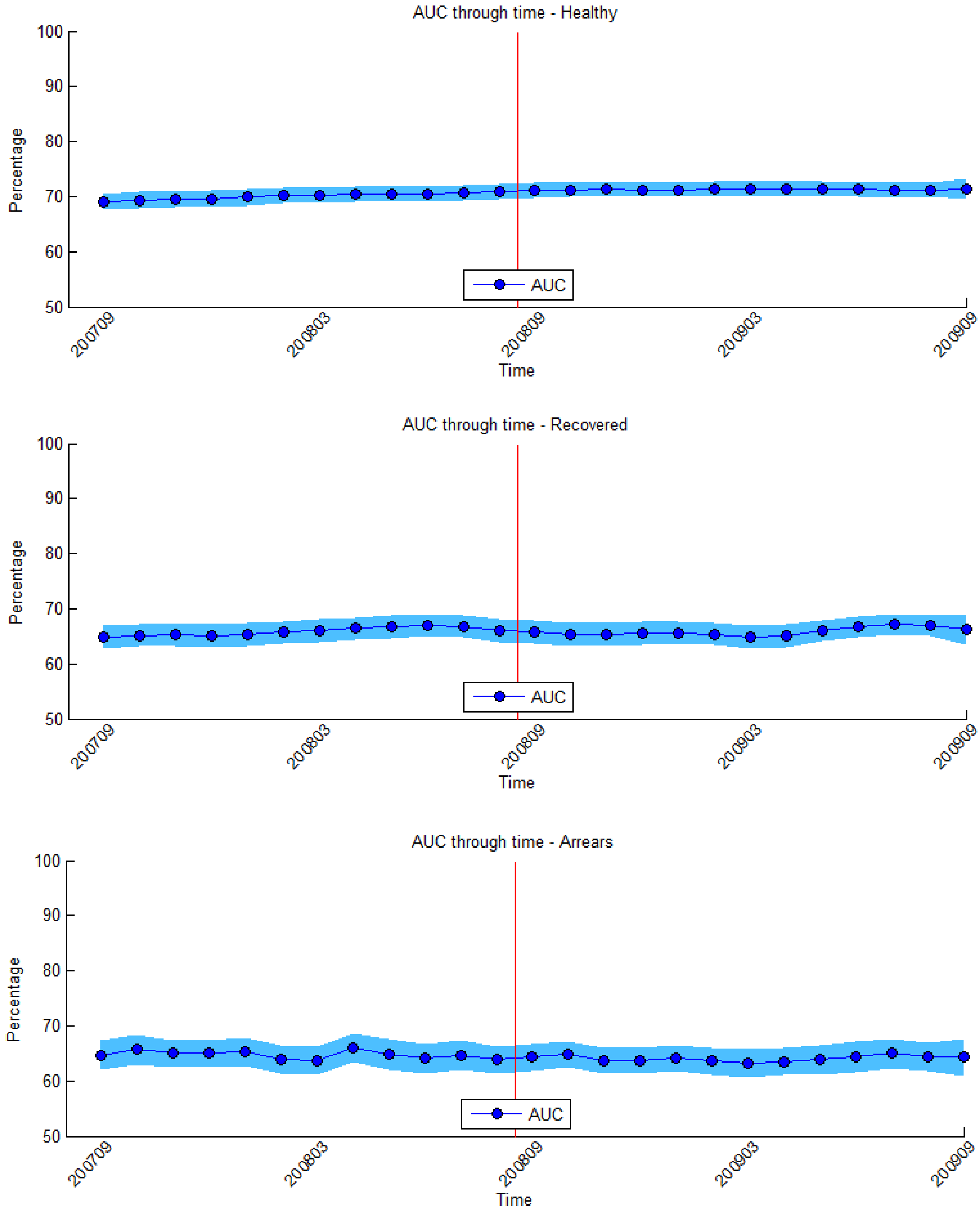


Figure 6-15 AUROC through time for sections 'Healthy', 'Recovered' and 'Arrears' - LT Logit prototype with Least Squares extrapolation

The Kolmogorov-Smirnov test shows that PDs of defaulted customers estimated by the LT Logit prototype with Least Squares extrapolation significantly differ from PDs estimated by the LT Logit prototype with Least Squares extrapolation of not defaulted customers. The test hypothesis is accepted for every moment of measurement.

Concluding it can be said that the LT Logit prototype with Least Squares extrapolation does not meet all the quantitative requirements. A material bias can be observed when the predictive power is visualized for the different sections. Also the AUROC is lower than the AUROC from the 1 year Logit prototype. However, it has become clear that the LT Logit prototype with Least Squares extrapolation performs better than the LT Logit prototype with mean Lambda extrapolation regarding both predictive power and discriminatory power.

Experiments with a direct 5 year Logit prototype showed a higher predictive power, based on a smoother fit of the PD and DR on the different sections, but a lower discriminatory power[31]. Based on these observations, one can assume that the LT Logit prototype with Least Squares extrapolation can be improved. Using more observations, generated by extra prototypes for multiple periods, can provide a better fit with the assumed cumulative Weibull distribution. It could also be the case that the Weibull distribution is a wrong assumption for one or multiple sections. One could try to work with different distributions for the different sections or to combine distributions for a better fit on a single section.

6.1.4 Extrapolation given 1 year PD

To gather meaningful insight in the predictive power of the LT Logit prototype given 1 year PD, the estimated PDs by the model are plotted against the realized DR on portfolio level in Figure 6.16. Important to notice is the vertical red line, which is the border of OOT and non OOT observations. From the figure can be concluded that the LT Logit prototype given 1 year PD estimations follows the realized DR. Besides that, no seasonal effects or constant bias can be observed in the estimations, whereby requirements 4 and 6[32] are met on portfolio level. The absence of seasonal effects is the result of the used sampling method, explained in chapter 5.

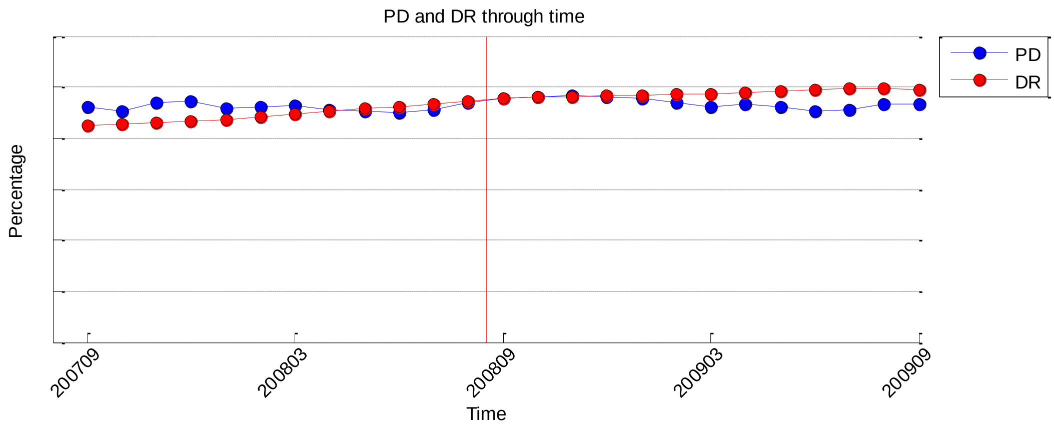


Figure 6-16 PD and DR through time - LT Logit prototype given 1 year PD

Figure 6.17 shows that the LT Logit prototype given 1 year PD fails in having a good predictive power. It performs worse than the other LT methods and a material bias of over- and under estimation, is visible for all the important sections $Healthy, Recovered$ and Arrears. Therefore the LT Logit Prototype given 1 year PD does not meet the requirements regarding predictive power on sectional level. The other sections, wherein the prototype performs better, can be found in APPENDIX 13. In this APPENDIX also visualized test results and visualized statistics, for example: the number of observations per section and moment of measurement, are presented.

[31] See APPENDIX 11 for the Logit direct 5 year prototype PD versus DR and AUROC figures.
[32] APPENDIX 2: Seasonal effects have to be observed and analysed.
APPENDIX 2: The model must have a good predictive power and no material deviation.

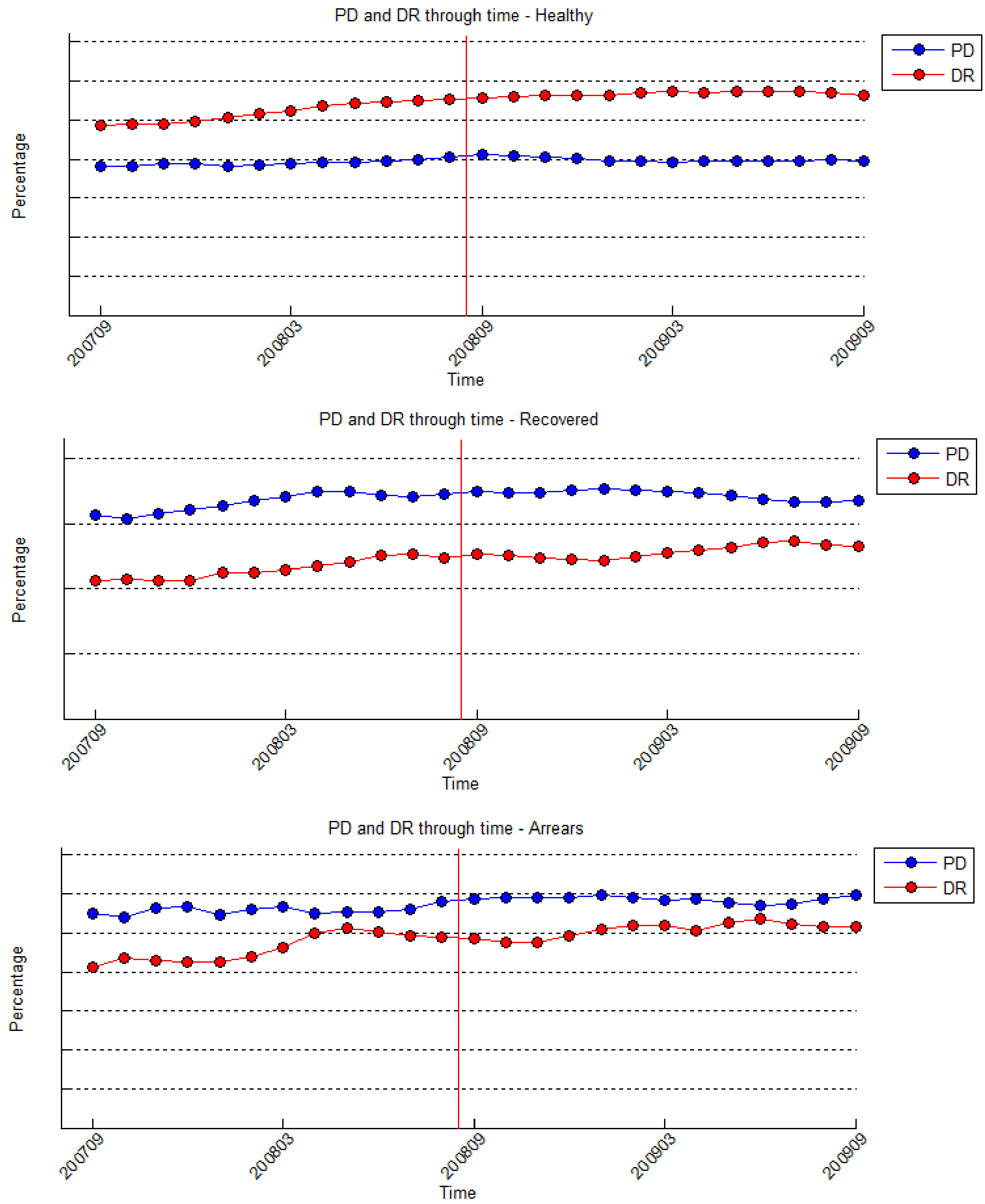


Figure 6-17 PD and DR through time for sections 'Healthy', 'Recovered' and 'Arrears' - LT Logit prototype given 1 year PD

Figure 6.18 show that the LT Logit prototype with Least Squares is also performs badly, whether or not passing the test, in comparison with the other two LT methods regarding the two sided binomial test on portfolio level, with a significance of $\alpha = 0.05$. The black cubes indicate that the realization lies inside the confidence interval, a white cube indicates that this is not the case. It is clearly that the result does not match with the expectation, that 95% of the realizations will lie in the confidence interval. This because the Logit prototype is not a purely TTC or PIT model, which results in failing the test as explained in paragraph 4.2.1. The results of the two sided binomial test on bucket level are presented in APPENDIX 13.

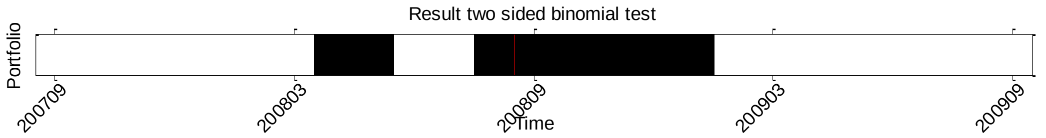


Figure 6-18 Result on portfolio level of the two sided binomial test- LT Logit prototype given 1 year PD

The Normal Test hypotheses are rejected for bucket 5 till 10. This could be the case because of the upward trend in DR, which results in OOT realizations which are all higher than the non OOT average DR. The hypotheses of the Vasicek test are accepted for all buckets on all moments of measurement.

To gather meaningful insight in the discriminatory power of the LT Logit prototype given 1 year PD, the AUROC is plotted a portfolio level in Figure 6.19. From the figure can be concluded that the Logit prototype with Least Squares extrapolation performs good regarding the discriminatory power. Because of this requirement 5 is met on portfolio level[33].

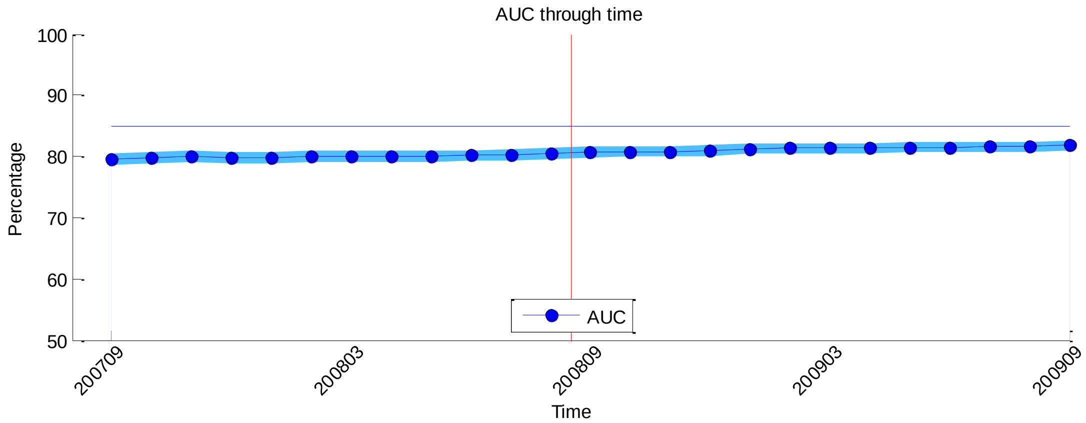


Figure 6-19 AUROC through time - LT Logit prototype given 1 year PD

Figure 6.20 shows that it is harder for the - LT Logit prototype with mean Lambda extrapolation prototype to succeed in having a good discriminatory power for different sections. In the Figure the important sections $Healthy, Recovered\ and\ Arrears$ are shown and the results are weak till moderate and quite constant. With those numbers the LT Logit prototype given 1 year PD is the worst performer regarding discriminatory power in comparison with the other LT methods. Because of this requirement 5 is not met on sectional level[34]. The AUROC figures for the other sections can be found in APPENDIX 13.

[33] APPENDIX 2: The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.
[34] See footnote 27.

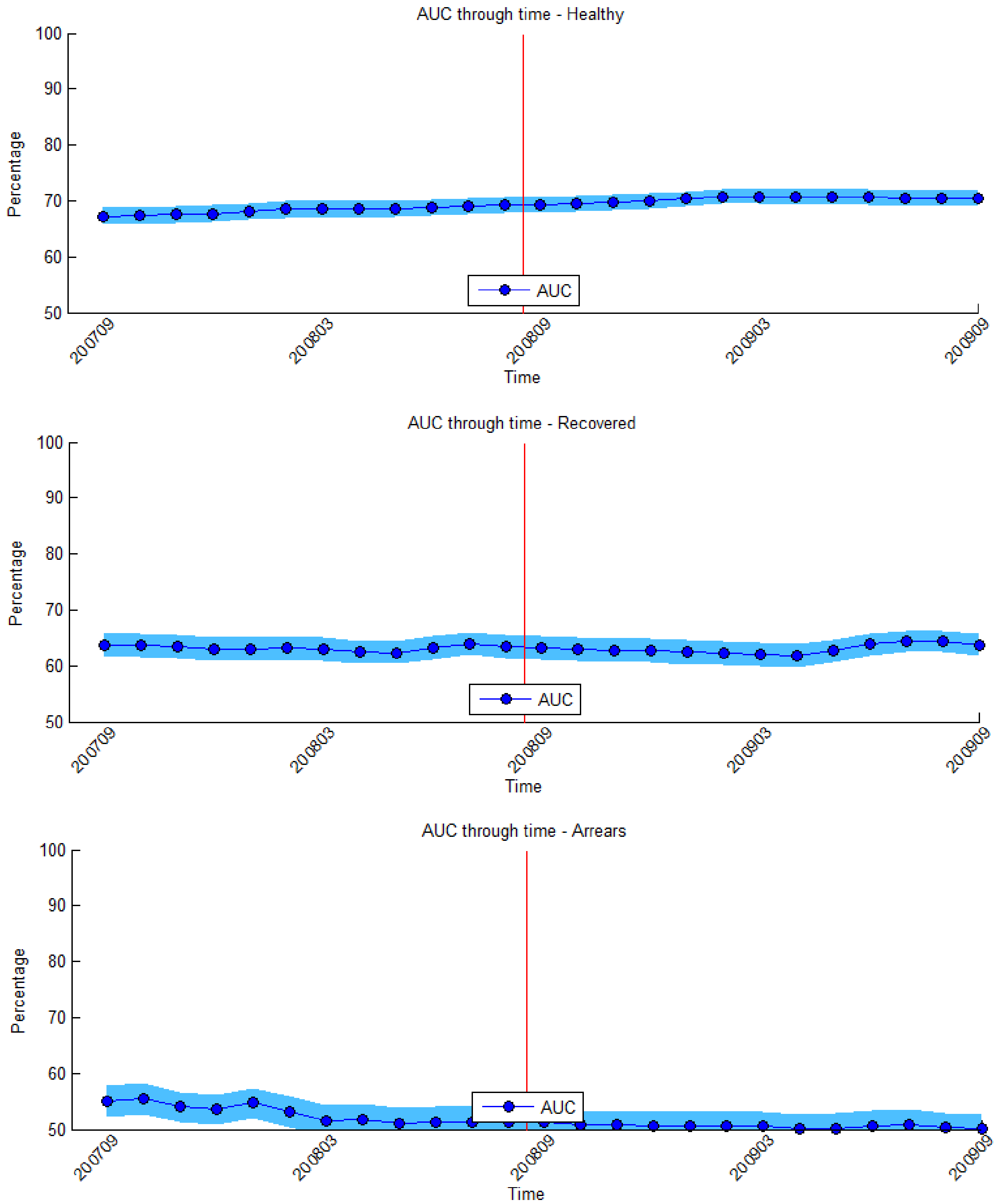


Figure 6-20 AUROC through time for sections 'Healthy', 'Recovered' and 'Arrears' - LT Logit prototype given 1 year PD

The Kolmogorov-Smirnov test shows that PDs of defaulted customers estimated by the LT Logit prototype given 1 year PD significantly differ from PDs estimated by the LT Logit prototype given 1 year PD of not defaulted customers. The test hypothesis is accepted for every moment of measurement.

Concluding can be said that LT Logit prototype given 1 year PD does not meet all the quantitative requirements. In comparison with the other LT methods, the LT Logit prototype given 1 year PD is the worst performer on both the aspects of predictive power and discriminatory power. The reason for this bad performance is probably that the PD for 1 year is not representative for going in default in the second, third, fourth of fifth year given a customer survived previous years. Despite the fact that the results of the LT Logit prototype given 1 year PD are not useful in this research, this finding can be useful in other research. It is known that numerous of other banks and consulting agencies are looking for a lifetime solution by estimating their 5 year PD based on solely their 1 year PD.

6.1.5 Summary

Summarizing it can be said that estimating LT PDs is more complex than estimating 1 year PDs, which matches with intuition. Nevertheless some of the methods used are capable of good estimation on portfolio level and reasonable estimating on sectional level. All the used LT methods guarantee that the PD rises if the period grows, for example: by assuming a cumulative distribution. This makes the gap between the 1 year PD and the LT PD explainable, so requirement 3 is met[35].

Tables 6.1 and 6.2 give an overview of the performance of the different methods used. As can be seen in Table 6.1 the 1 year Logit model has a good predictive power and a good discriminatory power and meets all the quantitative requirements. The LT Logit prototype with Least Squares extrapolation is the best LT method regarding predictive and discriminatory power. The LT Logit prototype with Least Squares extrapolation does not meet the requirements regarding predictive power on sectional level, but the predictive could be improved by performing the suggested improvements. This prototype will be used as LT prototype in the comparison of the frameworks.

6-1 Overview performance Logit 1 year prototype

Category	Test	Test Scope	Result
Predictive Power	Two sided Binomial test	Portfolio	Test failed for 50.68% of the moments of measurement
		Buckets	Test failed for 63.93% of the moments of measurement
	Normal test	Buckets	Test failed for buckets 4 till 10 (around 25% of total observations)
	Vasicek	Buckets	Test passed for all buckets on all moments of measurement
Discriminatory power	AUROC	Portfolio	88% (good)
		Healthy/Recovered/Arrears	70%/70%/71% (moderate - sufficient)
		NHG/non NHG	85%/88% (good)
		Eastern/Western Region	88%/88% (good)
	Kolmogorov-Smirnov	Portfolio	Test passed on all moments of measurement

[35] APPENDIX 3: The difference between a LT PD and a 12-months PD has to be explainable.

6-2 Overview performance different Logit LT methods

Category	Test	Test Scope	LT Logit prototype with mean Lambda extrapolation	LT Logit prototype with Least Squares extrapolation	LT Logit prototype given 1 year PD
Predictive Power	Two sided Binomial test	Portfolio	Test failed for 32.00% of the moments of measurement	Test failed for 12.00% of the moments of measurement	Test failed for 32.00% of the moments of measurement
		Buckets	Test failed for 48.33% of the moments of measurement	Test failed for 55.00% of the moments of measurement	Test failed for 53.00% of the moments of measurement
	Normal test	Buckets	Test failed for buckets 7 till 9 (around 53% of total observations)	Test failed for buckets 8 till 12 (around 42% of total observations)	Test failed for buckets 5-10 (around 89 % of total observations)
	Vasicek	Buckets	Test passed for all buckets on all moments of measurement	Test passed for all buckets on all moments of measurement	Test passed for all buckets on all moments of measurement
Discriminatory power	AUROC	Portfolio	81% (sufficient – good)	81% (sufficient – good)	81% (sufficient – good)
		Healthy/ Recovered/Arrears	70%/65%/52% (weak - sufficient)	70%/65%/65% (moderate - sufficient)	69%/63%/52% (weak - sufficient)
		NHG/ non NHG	70%/81% (sufficient – good)	80%/81% (sufficient – good)	76%/81% (sufficient – good)
		Eastern/ Western -Region	81%/81% (sufficient – good)	81%/80% (sufficient – good)	80%/80% (sufficient – good)
	Kolmogorov-Smirnov	Portfolio	Test passed on all moments of measurement	Test passed on all moments of measurement	Test passed on all moments of measurement

6.2 Logit framework - Qualitative Results

The performance of the Logit prototype on the 1 year results is comparable with the performance of the PHIRM 2.0 PD model, so requirement 5 is met[36]. The model is as result of the similarities with the PHIRM 2.0 PD model also explainable, whereby whish 14 is fulfilled[37]. The variables used in the Logit Prototype are selected by a heuristic, which does not ensure them to be intuitive. This makes that wish13[38] is not fulfilled, but this wish could be fulfilled if the variables will be selected with the help of experts. Furthermore are all the

[36] APPENDIX 3: The models don't have to perform better on the quantitative aspect than the currently used PHIRM 2.0 PD model, under the condition that the difference in quantitative performance is compensated on the qualitative aspect.
[37] APPENDIX 3: The used methods have to be explainable.
[38] APPENDIX 3: Used variables have to be intuitive.

decisions made and the assumptions assumed are documented in this report, whereby requirement 12 is met[39].

6.3 Tobit framework - Quantitative analysis & Results

As already mentioned in paragraph 4.2.2., the theoretical idea and the assumptions made during this research to use the Tobit framework for estimating PDs did not turn out to work in practice. In the first sub-paragraph the problem of PD estimation with the 1 year Tobit prototype is set out, because of this non-performance for a 1 year period, it has been decided not to develop Tobit prototypes for multiple periods. In the second sub-paragraph possible causes for the non-performance of the prototype are given. This paragraph ends with a summary.

6.3.1 The level problem

A good fit of the Tobit latent variable (y_i^*) could not be made, which resulted in a model with a far from sufficient predictive power. To illustrate this, the estimated PDs by the 1 year Tobit prototype are plotted against the realized DR on a portfolio level in Figure 6.21. Important to notice is the vertical red line, which is the border of OOT and non OOT observations. From the figure can be concluded that the 1 year Tobit prototype has a level problem. The PD line and the DR line have the same trend, but there is a more or less constant observable gap of around 1.7% on each moment of measurement.

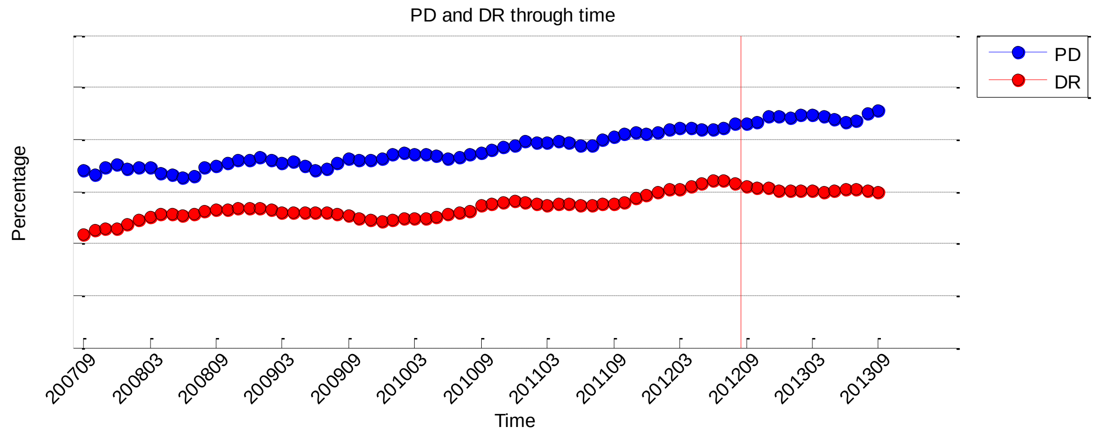


Figure 6-21 PD and DR through time on portfolio level – 1 year Tobit prototype

The level problem is a direct result of the lack of a decent fit, which is illustrated by an example. In Figure 6.22 the Loan To Foreclosure Value (LTFV) is plotted against the default score (both without outliers) for all observations of the Recovered customers. The LTFV variable is used in this example as the independent variable, because it normally has a considerable predictive power regarding dependent variables related to default events. So, if all assumptions made in this research are correct, one should expect a decent fit. As can be seen in the figure the latent Tobit variable ($y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_k$) does not lie in the middle of the observations. This should be case with a decent fit, as illustrated in the example in Figure 2.2. This observation indicates that at least one of the assumptions, or the application of the Tobit framework is wrong.

[39] APPENDIX 3: All choices and assumptions made have to be documented sufficiently.

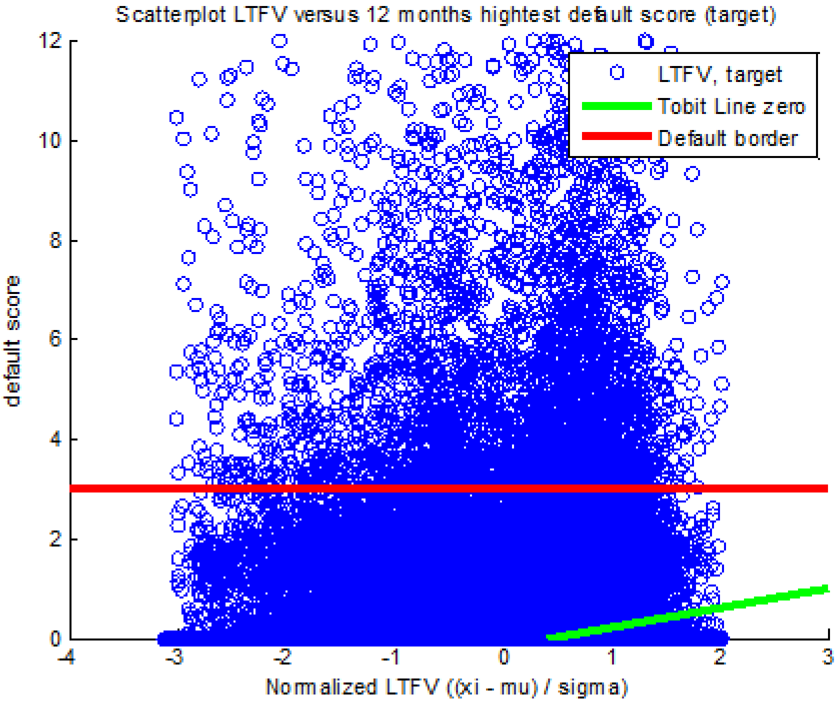


Figure 6-22 Loan To Foreclosure Value against default score

The fit illustrated in Figure 6.22 has as a result that the error terms are not normally distributed. This is illustrated in Figure 6.23. The histogram shown contains all observable errors $(y_i^* - y_i)$. So the errors from censored observations are not illustrated. If the errors were normally distributed the figure had to show the right half of a skewed normal curve starting at the x axis from zero, which is clearly not the case in this example.

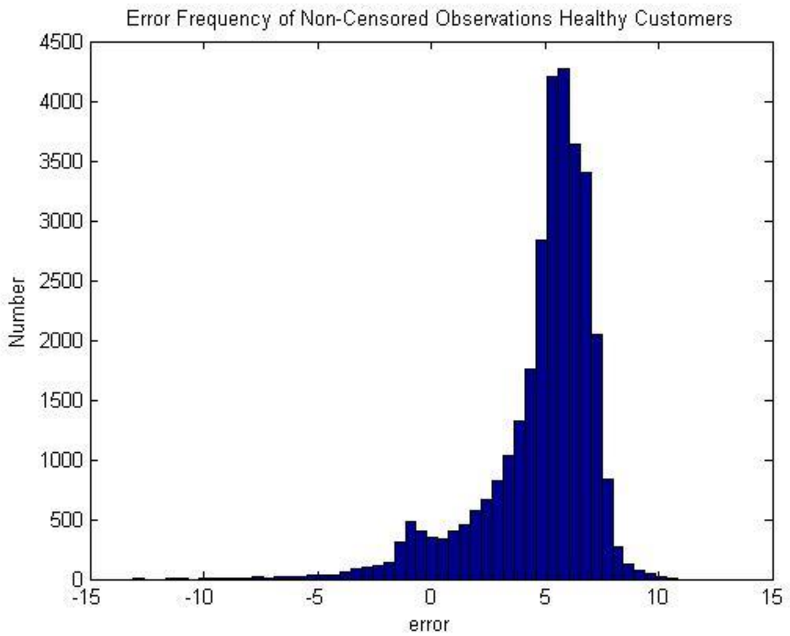


Figure 6-23 Histogram of observable error terms

Normally distributed errors are an important assumption in the theoretical idea of estimating PDs with the Tobit prototype. The sigma (σ) of the error distribution determines the level of the estimated PDs, as can be distracted from Equation 6-1. A higher sigma leads to a higher PD, and a lower sigma leads to a lower PD, so a wrong assumption about the sigma (caused by a bad fit) leads to a level problem.

$$PD_i = 1 - \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{3 - \mathrm{y}_\mathrm{i}^*}{\sigma\sqrt{2}}\right)\right], \qquad \textbf{(6-1)}$$

$y_i^* = estimated\ latent\ Tobit\ variable$

$\sigma = standard\ deviation\ of\ the\ normally\ distributed\ error\ term\ u_i$

Summarizing it can be said that the predictive power of the 1 year Tobit prototype is insufficient, because of a bad fit of the latent Tobit variable (y_i^*), which result in non-Normal distributed errors and a level problem. This does not mean that the Tobit framework is a non performing framework in itself. The bad fit could also be caused by wrong assumptions made in this research. The possible causes for the bad fit are studied in the next paragraph.

6.3.2 Possible causes

One of the possible causes for the bad fit, of the Tobit latent variable (y_i^*), could be the high proportion of censored observations. Defaults are rare events, and therefore a lot of observations are censored in this research. This lack of information could be too hard to handle for the Tobit framework, and maybe the Tobit framework should not have been used in this research. In table 6.3 the proportion of censored observations is set out for the sections $Healthy, Recovered\ and\ Arrears$. No further research is done regarding the effect of the proportion of censored observations on the quality of the fit, although it is an interesting topic, because the data used in PD estimation will always be characterized by a high level of censoring.

6-3 Proportion of censored observations per section ($\mathbf{Healthy, Recovered\ and\ Arrears}$)

Section	Percentage of censored observations
Healthy	96.45%
Recovered	81.41%
Arrears	60,31%

Another cause of the bad fit could be the distribution of the dependent variable (y_i) or default score. An important assumption in this research is that the depending variable has a continuous distribution. A continuously distributed depending variable is an important condition for using the Tobit framework. It could be the case that the depending variable used (y_i) is not continuous enough. In Figure 6.24 the value of the default scores for all non-censored observations are represented in a histogram. The peak at 0.0433 and the peak between 1 and 2 could be the reason for the dependent variable to be not continuous enough. The peaks are probably the result of the discrete character of the $\mathrm{ammount\ of\ arrears}$, which is an important part of the dependent variable $(\mathrm{ammount\ of\ arrears\ /\ monthly\ payment\ ammount})$. The dependent variable has a discrete character, because it does not happen often that a customer pays a part of the month term, instead of the whole term. Also business processes, like the handling of customers in arrears by the Arrears management department, could cause a non continues depending variable.

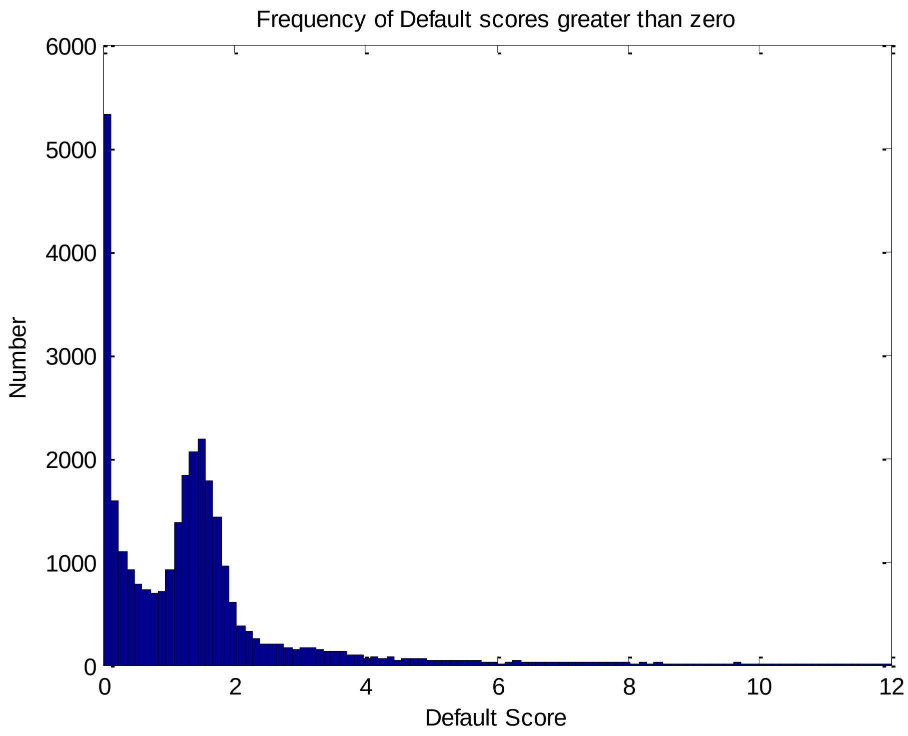


Figure 6-24 Number of default scores greater than zero

Another cause of the bad fit could be that all variables used during the estimation of the Tobit prototype have a lack of predictive power, which would result in a bad fit. This is not probable given the fact that the variables used work well in the Logit and Hazard prototypes. The reason that the fit is bad could also be a combination of multiple causes mentioned above.

6.3.3 Summary

Concluding it can be said that the theoretical idea and the assumptions made during this research to use the Tobit framework to estimate PDs did not work out as intended. The presence of a level problem as result of a bad fit, makes that the Tobit framework has an insufficient predictive power. Possible causes of the bad fit could be the high proportion of censored variables, the low predictive power of variables, or a wrong assumption about a not continuously distributed depending variable. Because the Tobit prototype is far from meeting the requirements and therefore could not be used for predicting PDs, the prototype will not be present during the comparison of the prototypes.

6.4 Tobit framework - Qualitative Results

Because the Tobit prototype and the related assumptions made during this research did not turn out to work, requirement 5 is not met[40]. The variables used in the Logit Prototype are selected by a heuristic, which does not ensure them to be intuitive. This makes that wish13[41] is not fulfilled, but this wish could be fulfilled if the variables will be selected with the help of experts. The Tobit framework used is relatively easy to interpret

[40] APPENDIX 3: The models don't have to perform better on the quantitative aspect than the currently used PHIRM 2.0 PD model, under the condition that the difference in quantitative performance is compensated on the qualitative aspect.
[41] APPENDIX 3: Used variables have to be intuitive.

and therefore whish 14 is fulfilled[42]. Furthermore are all the decisions made, assumptions assumed and directions for further research given are documented in this report, whereby requirement 12 is met[43].

6.5 Hazard framework - Quantitative analysis & Results

This paragraph consists out of 4 sub-paragraphs. The first sub-paragraph gives insight in the 1 year PD performance of the Hazard prototype. The second sub-paragraph gives insight in the LT PD performance of the same prototype. In the third sub-paragraph possible improvements of the Hazard prototype are given. This paragraph ends by summarizing the results. All tests are performed on an OOS dataset, covering both OOT and non OOT observations. The boundary of OOT and non OOT observations is August 2008.

6.5.1 1 year performance Hazard prototype

To gather meaningful insight in the predictive power of the Hazard prototype based on the 1 year results, the estimated PDs by the are plotted against the realized DR on a portfolio level in Figure 6.25. The Figure shows that the gap between the PD and DR increases as time passes. It is plausible to assume that this is caused by the way the estimation dataset is constructed, as explained in chapter 5, which could possibly be improved in further research. In this dataset most events $(default, prepayment, nothing\ observed)$ are linked with observations from September 2009, as illustrated in Figure 2.26. Because of this concentration of information around the beginning of the observation period, the prototype is better in estimating PDs at moments in time in the beginning of the observation period. This assumption is confirmed by the constant performance in the OOT period. No seasonal effects can be observed in the estimations, whereby requirement 4[44] is met. This is a result of the way the dataset is constructed, explained in chapter 5.

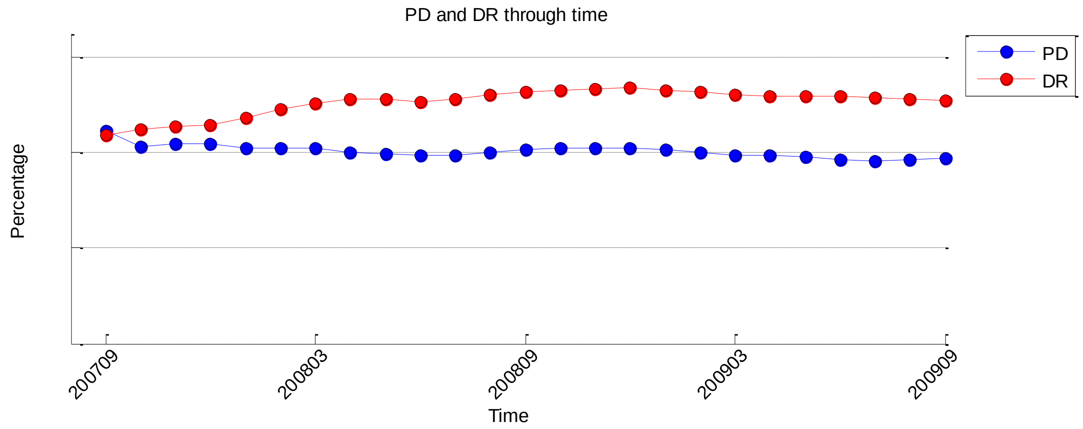


Figure 6-25 1 year PD and DR through time on portfolio level – Hazard prototype

[42] APPENDIX 3: The used methods have to be explainable.
[43] APPENDIX 3: All choices and assumptions made have to be documented sufficiently.
[44] APPENDIX 2: Seasonal effects have to be observed and analysed.

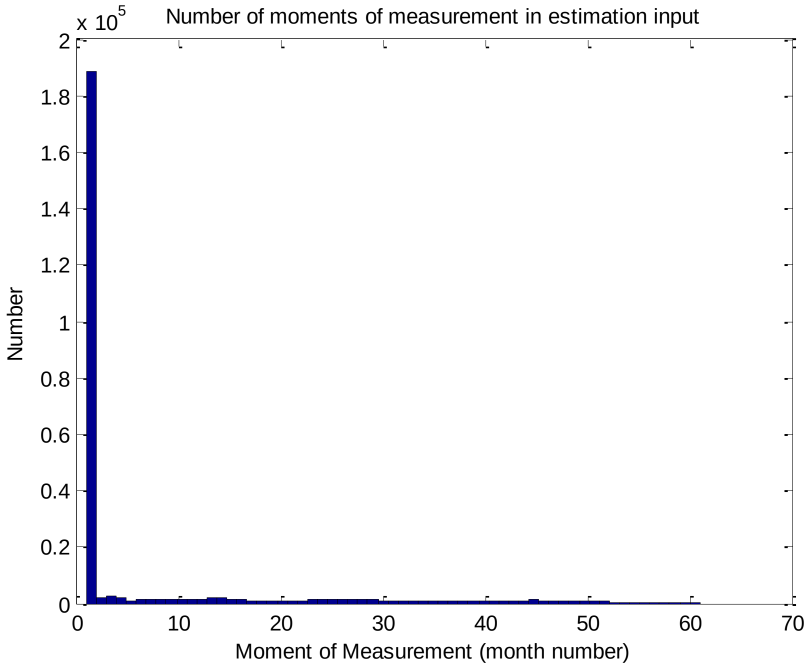


Figure 6-26 Number of moments of measurement in Hazard prototype estimation input

Figure 6.27 shows that the Logit 1 year prototype does not succeed in having a sufficient predictive power for different sections. In the Figure the important sections $Healthy, Recovered\ and\ Arrears$ are shown. The results for the other sections can be found in APPENDIX 9, together with visualized test results and visualized statistics, for example: the number of observations per section and moment of measurement. The 1 year predictive power of the Hazard prototype is better in the other sections than in the sections illustrated in Figure 6.27.

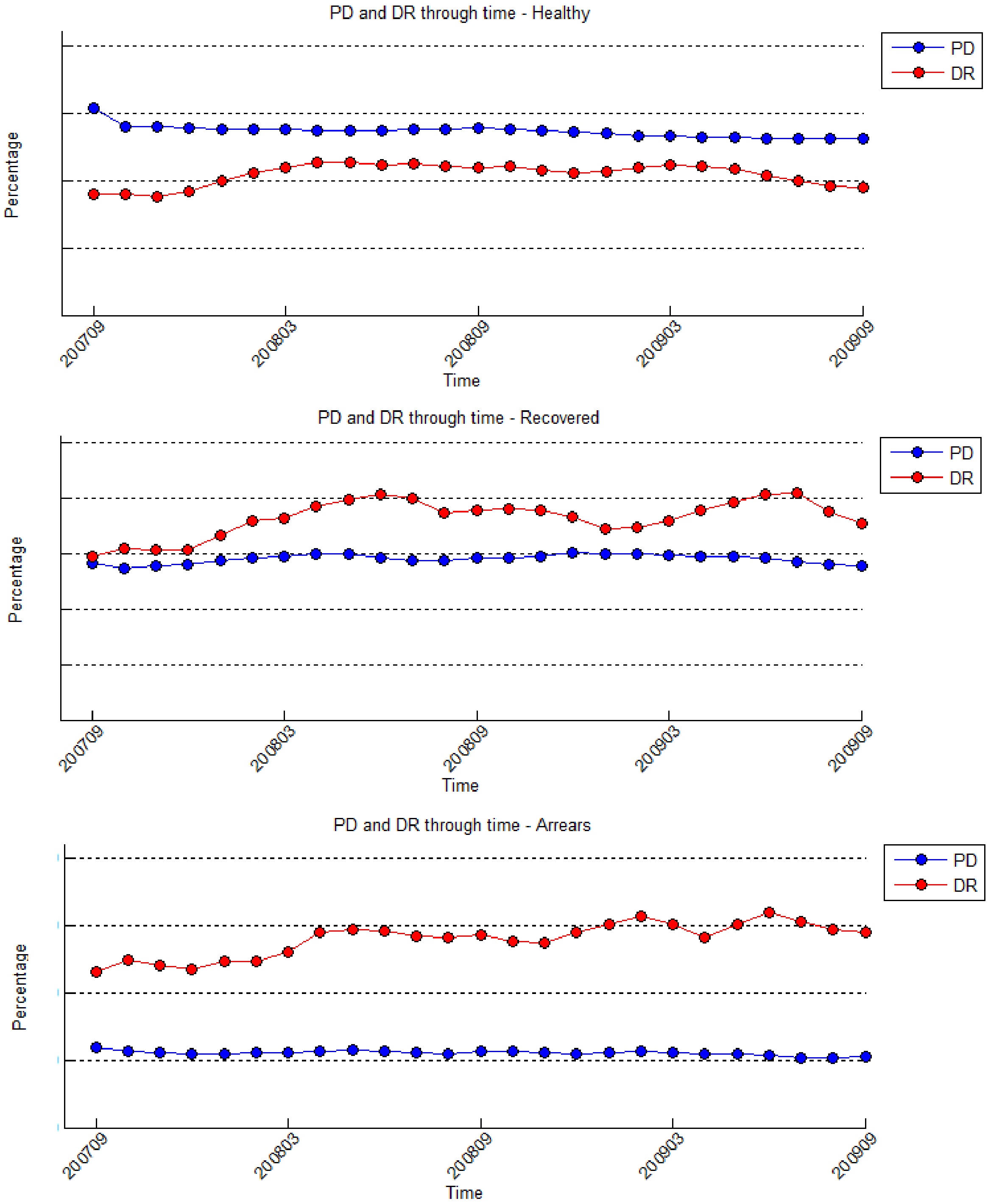


Figure 6-27 1 year PD and DR through time for sections 'Healthy', 'Recovered' and 'Arrears' – Hazard prototype

Figure 6.18 show that the hazard prototype fails the two sided binomial test on portfolio level, with a significance of $\alpha = 0.05$, for the majority of moments of measurement. The black cubes indicate that the realization lies inside the confidence interval, a white cube indicates that this is not the case. It is clearly that the result does not match with the expectation, that 95% of the realizations will lie in the confidence interval. This is mainly caused by the gap between PD and DR shown in Figure 6.25, but also because the Logit prototype is not a purely TTC or PIT model, which results in failing the test as explained in paragraph 4.2.1. The results of the two sided binomial test on bucket level are presented in APPENDIX 14.

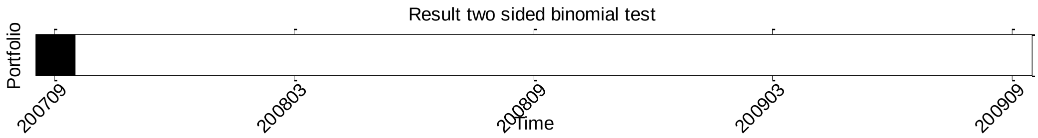


Figure 6-28 Two sided binomial test - Hazard prototype 1 year Result

The Normal Test hypotheses are rejected for bucket 1, 2, and 6 till 12. This could be the case because of the upward trend in DR, which results in OOT realizations which are all higher than the non OOT average DR. The hypotheses of the Vasicek test are rejected for buckets 9 till 11, probably caused by the gap between PD and DR shown in Figure 6.25.

To gather meaningful insight in the discriminatory power of the Hazard prototype based on the 1 year results, the AUROC is plotted on portfolio level in Figure 6.29. From the figure can be concluded that, based on 1 year performance, the Hazard prototype performs good regarding the discriminatory power. Because of this requirement 5 is met on portfolio level[45]. Notice that the AUROC is better in the first month of the observation period in comparison with the AUROC in the other months. This is the result of the dataset construction as explained during the discussion of the predictive power of the 1 year Hazard prototype.

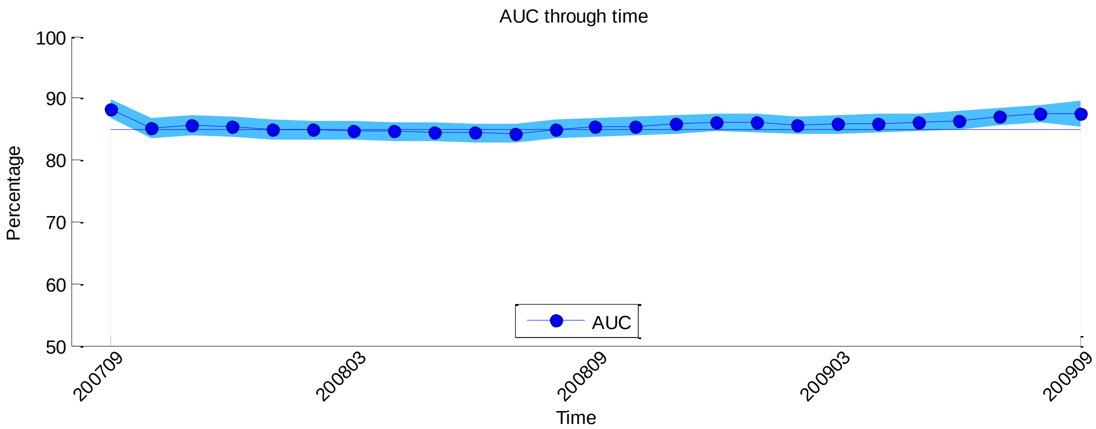


Figure 6-29 AUROC through time - Hazard prototype 1 year performance

Figure 6.30 shows that it is harder for the Hazard prototype to succeed in having a sufficient discriminatory power for different sections. In the Figure the important sections $Healthy, Recovered\ and\ Arrears$ are shown and the results are moderate till sufficient. With those numbers requirement 5 is not met on sectional level[46]. The AUROC figures for the other sections can be found in APPENDIX 14.

[45] APPENDIX 2: The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.
[46] See footnote 39.

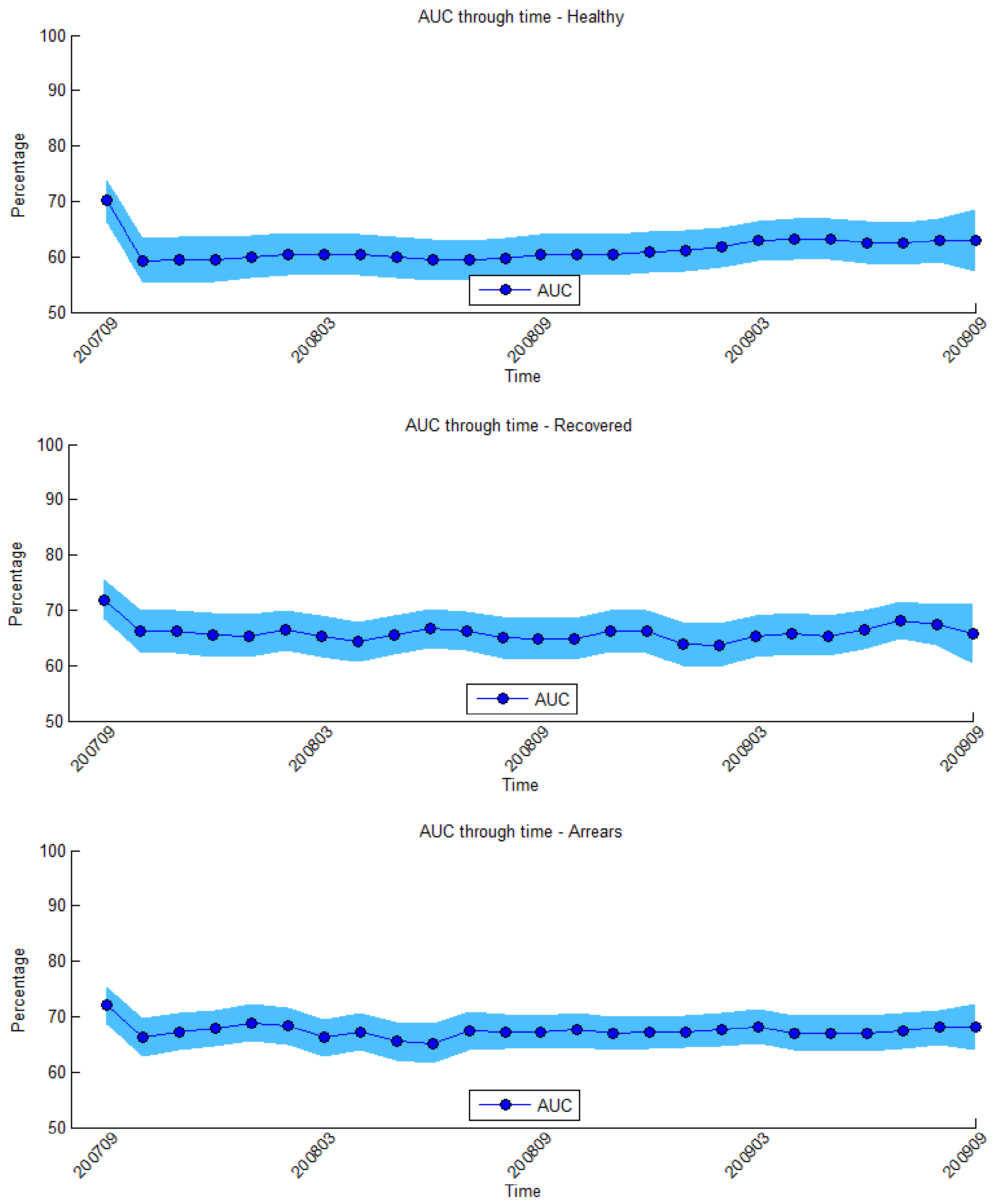


Figure 6-30 AUROC through time for sections 'Healthy', 'Recovered' and 'Arrears' – Hazard prototype 1 year performance

The Kolmogorov-Smirnov test shows that 1 year PDs of defaulted customers estimated by the Hazard prototype significantly differ from 1 year PDs estimated by the Hazard prototype of not defaulted customers. The test hypothesis is accepted for every moment of measurement.

Concluding can be said that, based on 1 year results, the Hazard prototype does not meet all the quantitative requirements. The prototype does not have a good predictive power but has a sufficient discriminatory power. However, there is room for improvement. The possible improvements will be presented in the third sub-paragraph.

6.5.2 LT performance Hazard prototype

To gather meaningful insight in the predictive power of the Hazard prototype based on the LT results, the estimated PDs by the model are plotted against the realized DR on portfolio level in Figure 6.31. The Figure shows that the gap between the PD and DR increases as time passes. The reasoning behind this increasing gap is similar as for the gap seen in the 1 year performance of the Hazard prototype. No seasonal effects can be observed in the estimations, whereby requirement 4[47] is met. This is, just as the increasing gap, the result of the way the dataset is constructed, explained in chapter 5.

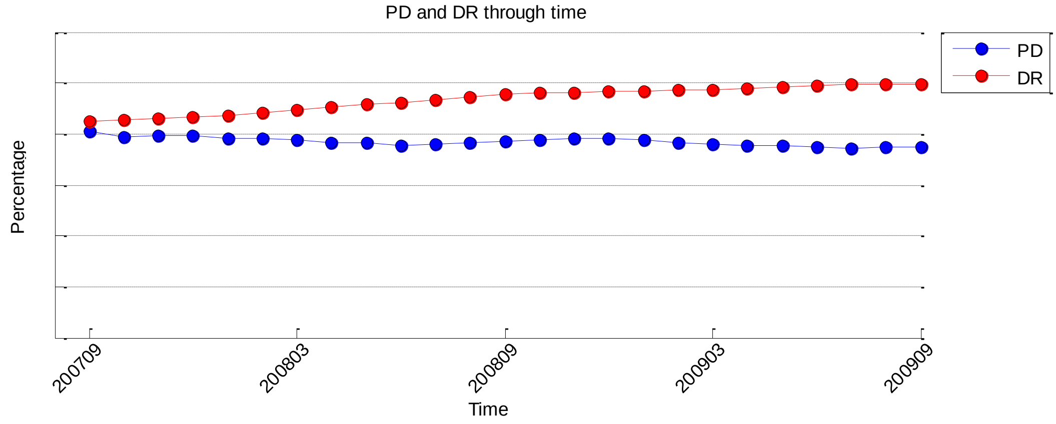


Figure 6-31 LT PD and DR through time on portfolio level – Hazard prototype

Figure 6.32 shows that the Logit 1 year prototype does not succeed in having a sufficient predictive power for different sections. In the Figure the important sections *Healthy, Recovered and Arrears* are shown. The LT results of the Hazard prototype on the sections show, just as the LT 1 year results of the Hazard prototype, the same remarkable appearance of being better in predicting PDs for Healthy customers than for Recovered customers and customers in Arrears. The reasoning behind this remarkable appearance is similar as for this appearance in the 1 year performance of the Hazard prototype.The figures of other sections, with a more equal spread of observations over the sections, can be found in APPENDIX 15. Also the visualized test results and visualized statistics, for example: the number of observations per section and moment of measurement, are presented in this APPENDIX.

[47] APPENDIX 2: Seasonal effects have to be observed and analysed.

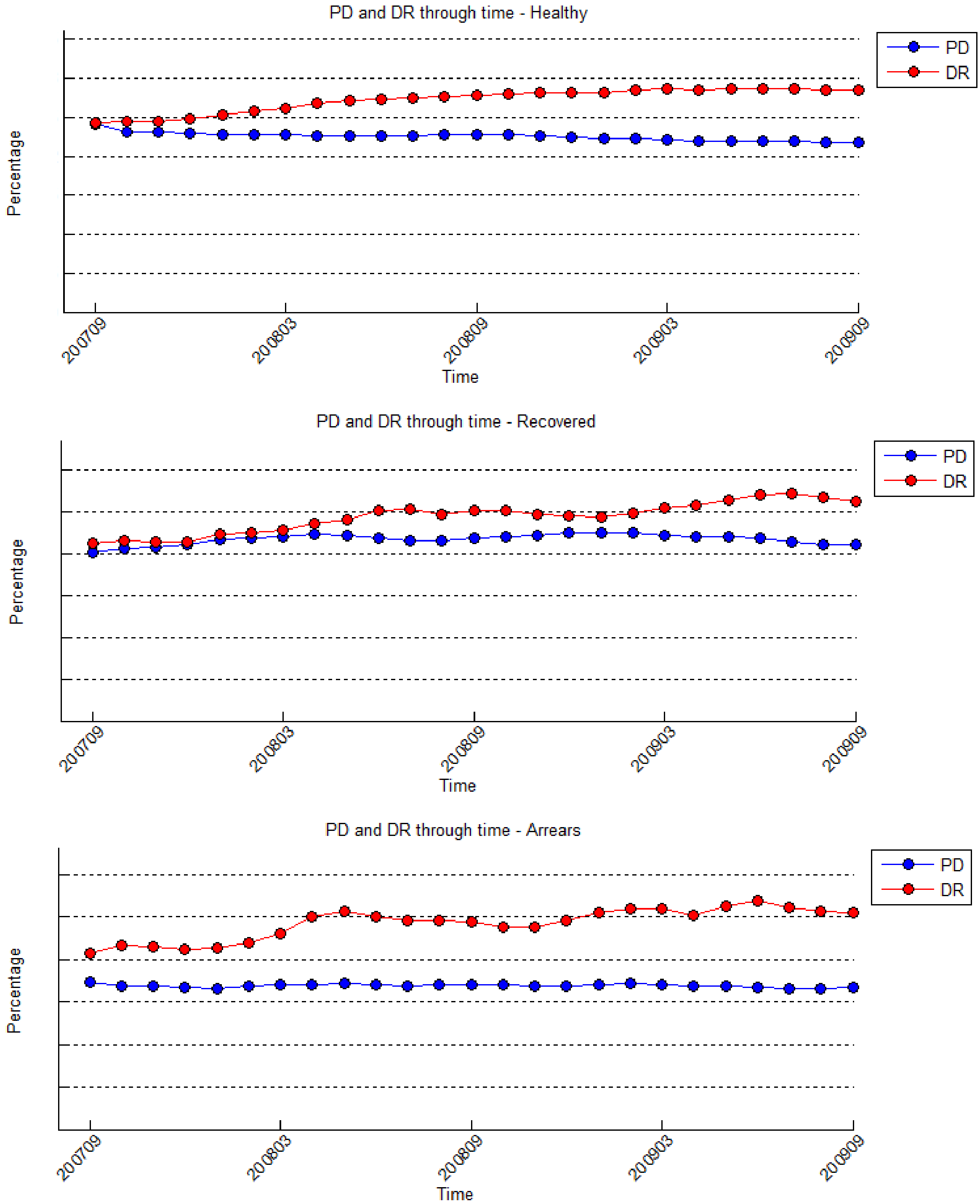


Figure 6-32 LT PD and DR through time for sections 'Healthy', 'Recovered' and 'Arrears' – Hazard prototype

Figure 6.33 shows that the hazard prototype fails to pass the two sided binomial test on portfolio level with a significance of $\alpha = 0.05$ for all the moments of measurement. The black cubes indicate that the realization lies inside the confidence interval, a white cube indicates that this is not the case. It is clearly that the result does not match with the expectation that 95% of the realizations will lie in the confidence interval. This is mainly caused by the gap between PD and DR in Figure 6.31, but also because the Logit prototype is not a purely TTC or PIT model, which results in failing the test as explained in paragraph 4.2.1. The results of the two sided binomial test on bucket level are presented in APPENDIX 15.

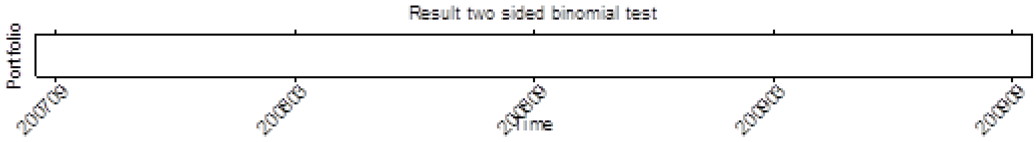


Figure 6-33 Two sided binomial test - Hazard prototype LT Result

The Normal Test hypotheses are rejected for buckets 5, 6 and 8 till 12. This could be the case because of the upward trend in DR, which results in OOT realizations which are all higher than the non OOT average DR. The hypotheses of the Vasicek test is rejected for buckets 5 and 9.

To gather meaningful insight in the discriminatory power of the Hazard prototype based on the LT results, the AUROC is plotted a portfolio level in Figure 6.34. From the figure can be concluded that, based on LT results, the Hazard prototype performs sufficient till good regarding the discriminatory power. Because of this, requirement 5 is met on portfolio level[48]. Notice that the AUROC is better in the first month of the observation period in comparison with the other months. This is the result of the dataset construction as also explained during the discussion of the 1 year predictive power of the Hazard prototype.

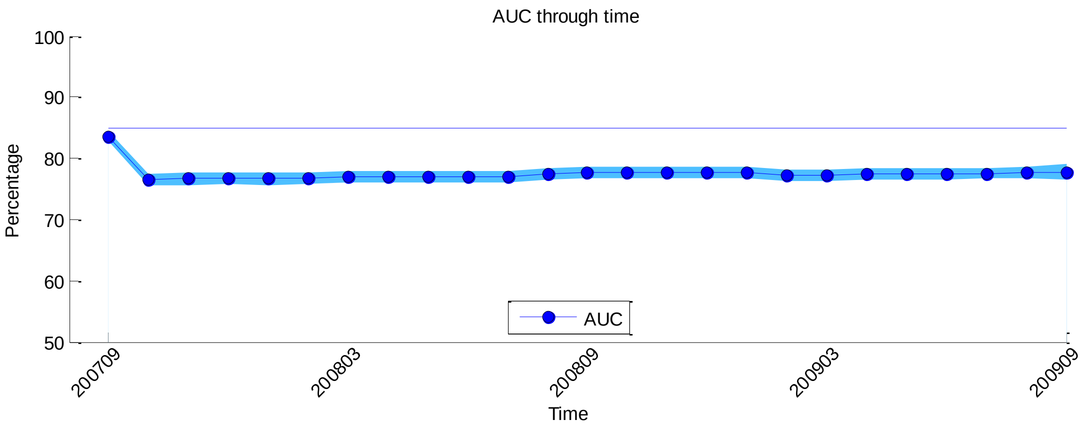


Figure 6-34 AUROC through time - Hazard prototype LT performance

Figure 6.35 shows that it is harder for the Hazard prototype to succeed in having a sufficient discriminatory power for different sections. In the Figure the important sections $Healthy, Recovered\ and\ Arrears$ are shown and the results are moderate till sufficient. With those numbers is requirement 5 is not met on sectional level[49]. The AUROC figures for the other sections can be found in APPENDIX 15.

[48] APPENDIX 2: The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.
[49] See footnote 42.

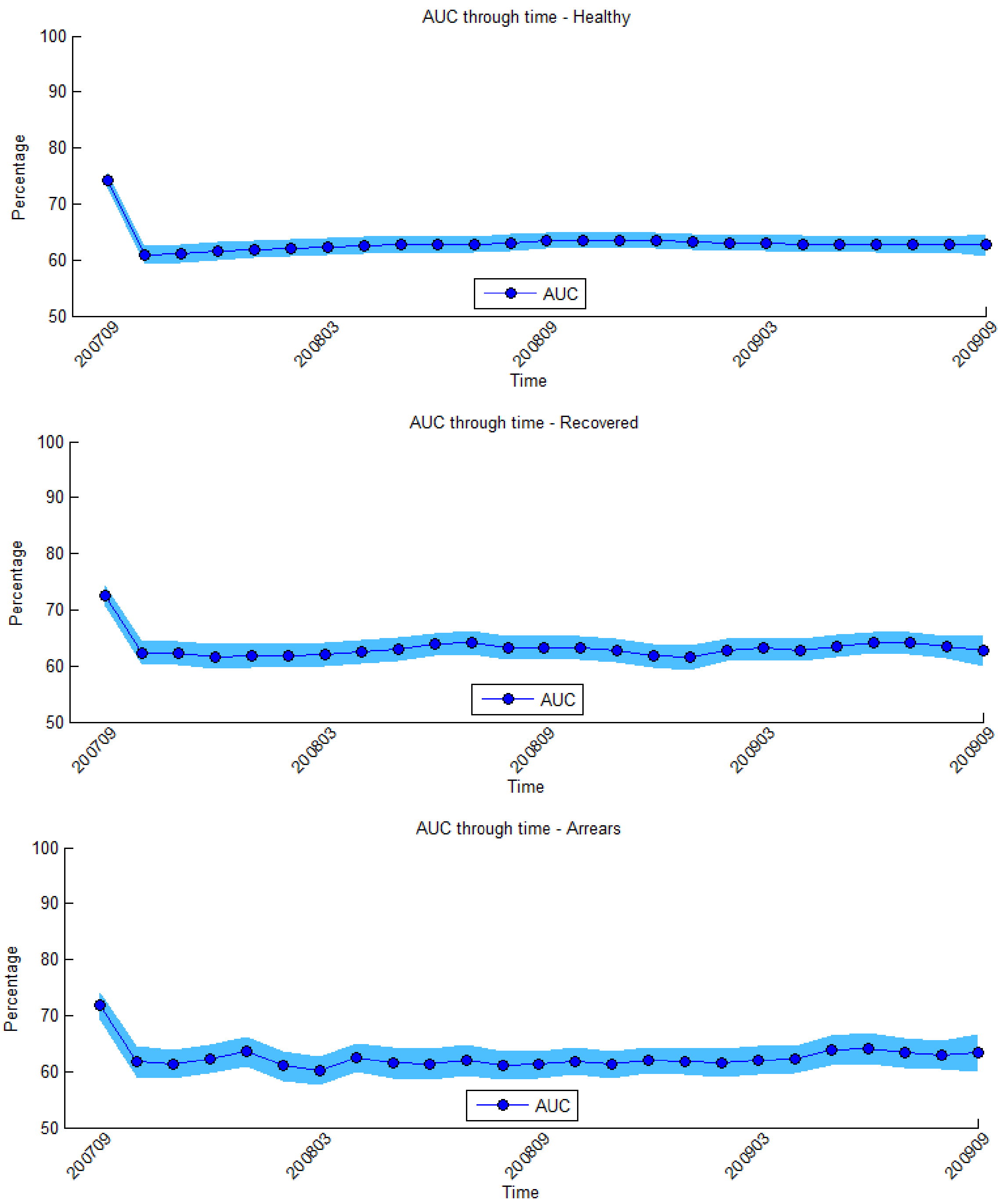


Figure 6-35 AUROC through time for sections 'Healthy', 'Recovered' and 'Arrears' – Hazard prototype LT performance

The Kolmogorov-Smirnov test shows that LT PDs of defaulted customers estimated by the Hazard prototype significantly differ from LT PDs estimated by the Hazard prototype of not defaulted customers. The test hypothesis is accepted for every moment of measurement.

Concluding can be said that the LT performance of the Hazard prototype is just as the 1 year performance not sufficient enough to meet all the quantitative requirements. The prototype does not have a good predictive power and a sufficient discriminatory power. However, there is room for improvement. In the next sub-paragraph possible improvements will be presented.

6.5.3 Possible improvements

The test results showed that the Hazard prototype does not have the predictive power aimed for. One of the reasons for this lack of performance could be an insufficient fit. To discover if this is the case, the fit of the Survival curve, the hazard rate of Default and the hazard rate of prepayment are plotted against the realizations. These three components together form the Hazard formula. A bad fit of one or multiple of these three components could cause a lack of predictive power.

In Figure 6.36 the fit of the Survival Curve is plotted against the realizations in the months of 2007. As can be seen in the figure, the fit is not as good as expected. One should expect the fit to be somewhere in the middle of the realizations. The fit is too low between 20 and 40 months and to high after 70 months. When this trend continuous the gap between realizations and the fit will increase as the time increases a well. The Survival Curve consists out of an exponential combination of both the hazard rate for default and the hazard rate for Prepayment. The lousy fit suggests that at least one of the hazard rates haves an insufficient fit.

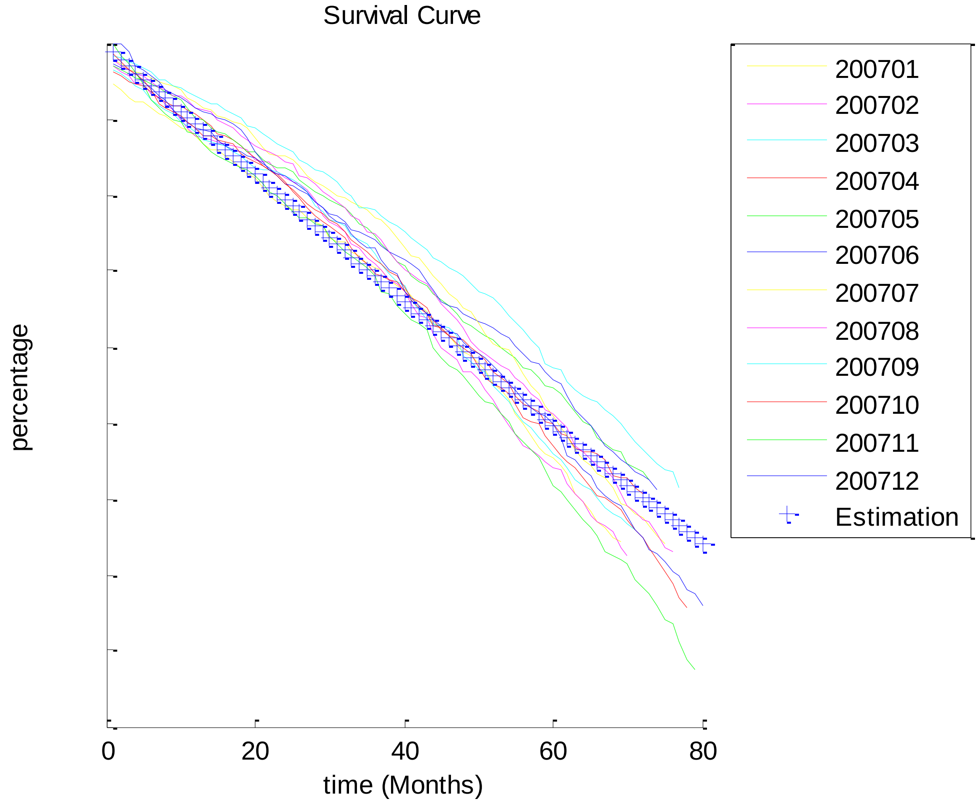


Figure 6-36 Fit of Hazard Survival curve

In Figure 3.37 the hazard rate of Prepayment is plotted against the realizations in the months of 2007. The figure is also a cumulative plot, because this is easier to interpret. The figure does not give evidence of a bad fit of the hazard rate of Prepayment and the realizations.

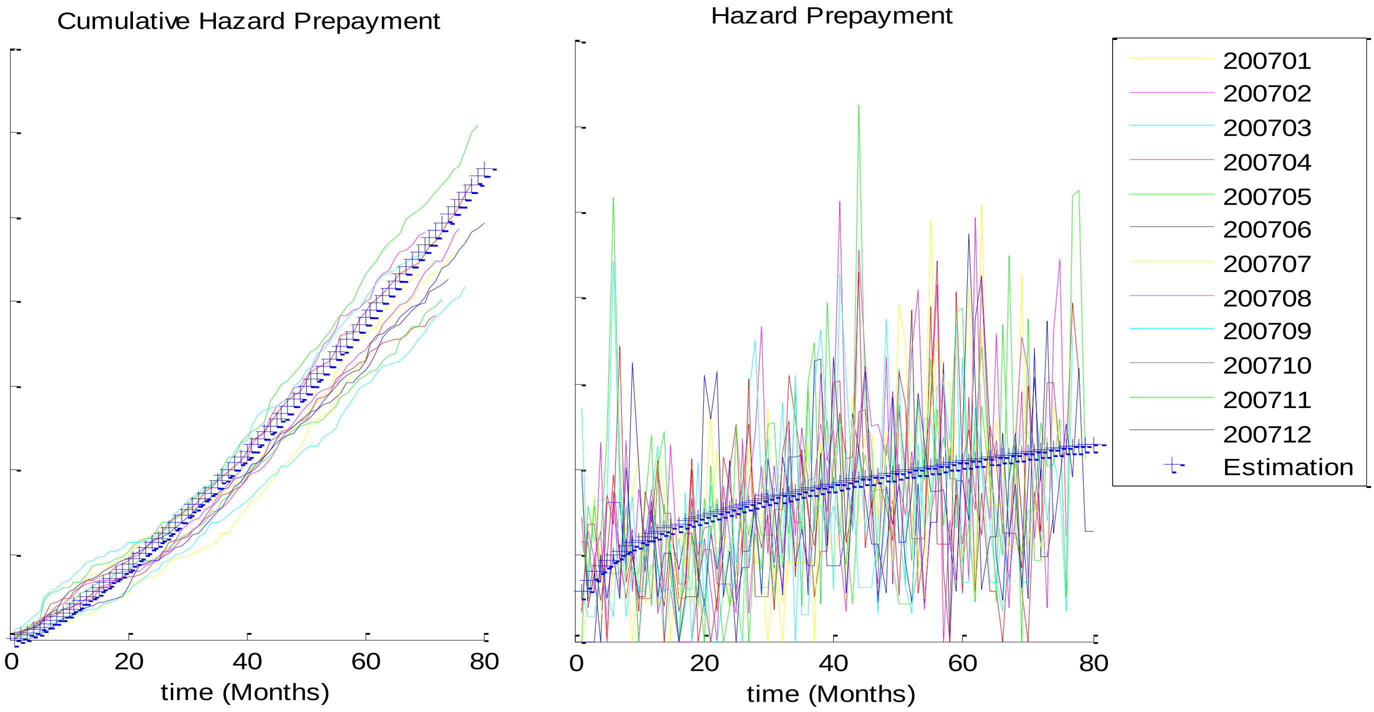


Figure 6-37 Fit of Hazard Prepayment and Cumulative Hazard Prepayment

In Figure 3.38 the hazard of Default is plotted against the realizations in the months of 2007. The figure contains also a cumulative plot, because this plot is easier to interpret. The figure does give evidence of a bad fit of the hazard rate of Default and the realizations. The estimated fit does lie in the middle of the observations, as expected in the case of a sufficient fit.

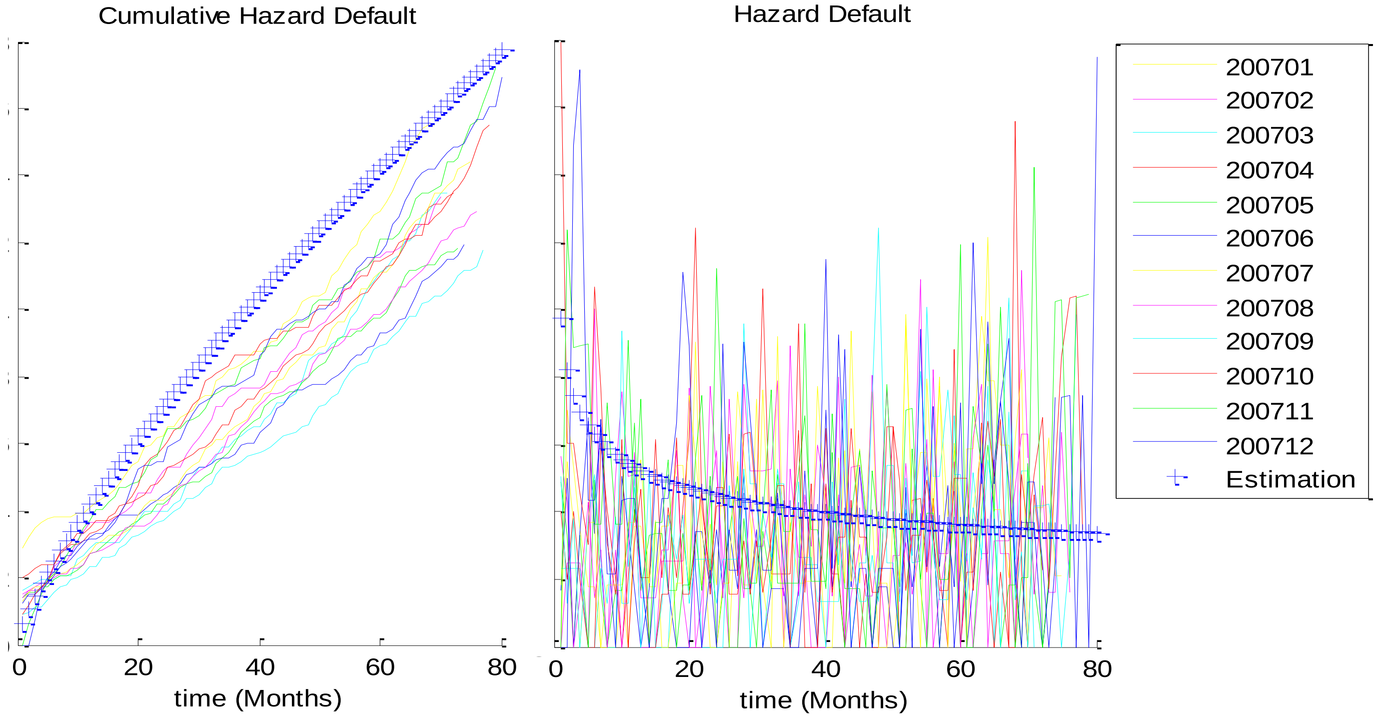


Figure 6-38 Figure 6 36 Fit of Hazard Default and Cumulative Hazard Default

The hazard prototype could be improved by providing a better fit of the hazard rate of Default. A possible direction of solution is to assume another distribution of Defaults. A possible distribution to try is the Log-logistic distribution, because this distribution could be used for events with first an increasing frequency followed by a decreasing frequency as observed in default figures (Bennett, 1993).

Other possible improvements of the performance of the Hazard Prototype are related to the way the dataset is constructed. As discussed in paragraph 6.3.1.1. the current way of constructing the dataset causes a bias in the input for estimating the Hazard prototype, regarding the spread of observations over the moments of measurements. A direction of solution for this problem could be the creation of more observations starting from some random moments during the original observation period per customer. This solution also helps to reduce the wryness regarding to the spread of observations over the different sections. It could also be tried to remove the separation of customers over the sections $(Healthy, Recovered, Arrears)$. It might be possible that a single model, estimated with more observations, is able to predict with a higher predictive power. Important to notice is that for the current way of constructing the dataset, it is assumed that defaults could be reset, in the sense that after a default a new observation could start. If this assumption does not hold, a new way of selecting data has to be conceived.

The Hazard prototype could possibly also be improved by performing a more extensive variable selection. Making the CRIME available for dealing with Hazard variables will ensure that the variables with the highest predictive power will be selected, which is not the case now, because of the used heuristic.

6.5.4 Summary

Summarizing it can be said that the Hazard prototype does not meet the requirements regarding predictive power. This is probably caused by the bad fit of the hazard rate of Default with the realizations. This problem could be resolved as explained in paragraph 6.3.1.3. Furthermore improvements in the construction of the dataset and the selection of variables could be made. The possible improvements, together with the fact that the Hazard framework offers a natural LT framework in which two events could be fitted, makes the Hazard prototype a promising prototype. The Hazard framework guarantees that the PD rises if the period grows, by assuming a cumulative distribution. Therefore the gap between the 1 year PD and the LT PD is explainable, so requirement 3 is met[50].

Tables 6.1 and 6.2 gives an overview of the 1 year and LT performance of the Hazard prototype.

[50] APPENDIX 3: The difference between a LT PD and a 12-months PD has to be explainable.

6-4 Overview 1 year performance Hazard Prototype

Category	Test	Test Scope	Result
Predictive Power	Two sided Binomial test	Portfolio	Test failed for 0.40% of the moments of measurement
		Buckets	Test failed for 43.33% of the moments of measurement
	Normal test	Buckets	Test failed for buckets till 1, 2, 6 till 12 (around 9% of total observations)
	Vasicek	Buckets	Test failed for all buckets 9 till 11 (around 4% of total observations)
Discriminatory power	AUROC	Portfolio	86% (good)
		Healthy/Recovered/Arrears	60%/65%/65% (moderate)
		NHG/non NHG	85%/86% (good)
		Eastern/Western Region	87%/87% (good)
	Kolmogorov-Smirnov	Portfolio	Test passed on all moments of measurement

6-5 Overview LT performance Hazard Prototype

Category	Test	Test Scope	Result
Predictive Power	Two sided Binomial test	Portfolio	Test failed for 0.00% of the moments of measurement
		Buckets	Test failed for 46.33% of the moments of measurement
	Normal test	Buckets	Test failed for buckets 5,6 and 8 till 12 (around 27% of total observations)
	Vasicek	Buckets	Test failed for all buckets 5 and 9 (around 0.1% of total observations)
Discriminatory power	AUROC	Portfolio	76% (sufficient)
		Healthy/Recovered/Arrears	60%/62%/62% (moderate)
		NHG/non NHG	70%/78% (sufficient)
		Eastern/Western Region	77%/77% (sufficient)
	Kolmogorov-Smirnov	Portfolio	Test passed on all moments of measurement

6.6 Hazard framework - Qualitative Results

The performance of the Hazard prototype on the 1 year results is less than the performance of the PHIRM 2.0 PD model, so requirement 5 is not met[51]. Notice that there are possibilities to improve this performance.

[51] APPENDIX 3: The models don't have to perform better on the quantitative aspect than the currently used PHIRM 2.0 PD model, under the condition that the difference in quantitative performance is compensated on the qualitative aspect.

The hazard framework offers a natural framework for estimating PDs, whereby whish 14 is fulfilled[52]. The variables used in the Hazard Prototype are selected by a heuristic, which does not ensure them to be intuitive. This makes that wish 13[53] is not fulfilled, but this wish could be fulfilled if the variables will be selected with the help of experts. Furthermore are all the decisions made and assumptions assumed are documented in this report, whereby requirement 12 is met[54].

6.7 Comparison of the Logit and Hazard prototypes

Tables 6.7 and 6.8 gives an overview of the performance of the Logit and Hazard prototypes for a 1 year and LT period. The Tobit prototype is not present in the tables, because the theoretical idea of using the Tobit framework to estimate PDs failed to work in practice.

As previously concluded is the Logit prototype the only prototype that meets almost all quantitative requirements. The Logit prototype has a good predictive- and discriminatory power on portfolio level and a sufficient till good predictive- and discriminatory power on sectional level, after performing improvements of the LT prototype. The Logit prototype also meets almost all the qualitative requirements. The only wish that is not fulfilled yet[55], could be fulfilled by selecting the variables with the help of experts.

Despite the Logit prototype outperforming the Hazard prototype, the Hazard prototype has potential. The Hazard model does not have a sufficient predictive power, but this could be resolved by making some improvements. If a Hazard prototype could be developed with the same or a better performance than the Logit prototype, this Hazard prototype would be preferred because it would offer a natural framework (no extrapolation is needed), which eases demonstrations of the prototype and explaining the gap between 1 year PDs and LT PDs. All qualitative requirements could be fulfilled in a, to be developed, improved Hazard prototype.

[52] APPENDIX 3: The used methods have to be explainable.
[53] APPENDIX 3: Used variables have to be intuitive.
[54] APPENDIX 3: All choices and assumptions made have to be documented sufficiently.
[55] APPENDIX 3: Used variables have to be intuitive.

6-6 Overview 1 year PD performance of the different Frameworks

Category	Test	Test Scope	Logit 1 year prototype	Hazard prototype 1 year
Predictive Power	Two sided Binomial	Portfolio	Test failed for 50.68% of the moments of measurement	Test failed for 0.40% of the moments of measurement
		Buckets	Test failed for 63.93% of the moments of measurement	Test failed for 43.33% of the moments of measurement
	Normal test	Buckets	Test failed for buckets 4 till 10 (around 25% of total observations)	Test failed for buckets till 1, 2, 6 till 12 (around 9% of total observations)
	Vasicek	Buckets	Test passed for all buckets on all moments of measurement	Test failed for buckets 9 till 11 (around 4% of total observations)
Discriminatory power	AUROC	Portfolio	88% (good)	86% (good)
		Healthy/ Recovered/Arrears	70%/70%/71% (moderate - sufficient)	60%/65%/65% (moderate)
		NHG/ non NHG	85%/88% (good)	85%/86% (good)
		Eastern/ Western -Region	88%/88% (good)	87%/87% (good)
	Kolmogorov-Smirnov	Portfolio	Test passed on all moments of measurement	Test passed on all moments of measurement

6-7 Overview LT PD performance of the different Frameworks

Category	Test	Test Scope	LT Logit prototype with Least Squares extrapolation	Hazard prototype LT
Predictive Power	Two sided Binomial	Portfolio	Test failed for 12.00% of the moments of measurement	Test failed for 0.00% of the moments of measurement
		Buckets	Test failed for 55.00% of the moments of measurement	Test failed for 46.33% of the moments of measurement
	Normal test	Buckets	Test failed for buckets 8 till 12 (around 42% of total observations)	Test failed for buckets 5,6 and 8 till 12 (around 27% of total observations)
	Vasicek	Buckets	Test passed for all buckets on all moments of measurement	Test failed for all buckets 5 and 9 (around 0.1% of total observations)
Discriminatory power	AUROC	Portfolio	81% (sufficient – good)	76% (sufficient)
		Healthy/ Recovered/Arrears	70%/65%/65% (moderate - sufficient)	60%/62%/62% (moderate)
		NHG/ non NHG	80%/81% (sufficient – good)	70%/78% (sufficient)
		Eastern/ Western -Region	81%/80% (sufficient – good)	77%/77% (sufficient)
	Kolmogorov-Smirnov	Portfolio	Test passed on all moments of measurement	Test passed on all moments of measurement

7 Conclusion and Recommendations

During this Thesis three prototypes are developed to estimate PDs, with three different statistical frameworks, under the conditions given by model owners and regulators. It turned out that one of the prototypes, based on the Tobit framework and the assumptions made during this research, was not able to sufficiently predict PDs, because of a lack of predictive power. This lack of predictive power is caused by the presence of a level problem as result of a bad fit of the latent Tobit variable. Possible causes of this bad fit could be the high proportion of censored variables, a to low predictive power of the variables or a wrong assumption about a not continuously distributed depending variable.

The best performing prototypes are based on the Logit framework, the framework that is currently used in the PHIRM 2.0 PD model. The 1 year Logit prototype has a good predictive power and also a good discriminatory power and meets all the quantitative requirements. All qualitative requirements could be met by selecting variables in consultation with experts. The LT Logit prototype with Least Squares extrapolation is the best LT method regarding predictive and discriminatory power. It does not meet the requirement regarding predictive power on sections. This could be improved by using more observations, generated by extra prototypes for multiple periods or the choice for another or combined distribution. The results show that predicting LT PDs is more complex than estimating 1 year PDs, which matches with intuition.

The third prototype developed is based on the Hazard framework, which offers a natural framework (no extrapolation is needed) for predicting life time events. The Hazard prototype does not meet the requirements regarding predictive power. This is probably caused by the bad fit of the hazard of Default and the realizations. This problem could be resolved, by the choice for another or combined distribution, and also improvements in the construction of the dataset and the selection of variables could be made. When these problems are resolved the predictive- and discriminatory power are expected to grow considerably, which makes the Hazard prototype a promising prototype. If a Hazard prototype could be developed with the same or a better performance than the Logit prototype, this Hazard prototype would be preferred because it is a natural framework, which eases demonstrations of the prototype and explaining the gap between 1 year PDs and LT PDs.

Given this conclusion the SNS Bank N.V. is advised to use the 1 year Logit prototype and the LT Logit prototype with Least Squares extrapolation (with improvements) as champion models and the Hazard prototype as challenger model. In this setup the Logit prototypes will be used as ‘current used’ models and the Hazard prototype can be improved via the possible improvements while being live (R. Chu, 2007).

8 Discussion & Suggestions for Further Research

Modelling is all about creating the best possible representation of reality. Therefore, this discussion is mainly about the gap between the developed prototypes and reality. First some general remarks are given, which are applicable for all the frameworks, complemented with suggestions for further research. After that, remarks are given per prototype, which has some overlap with the possible improvements mentioned in chapter 6. The prototype related remarks are also complemented with suggestions for further research.

In this report it is assumed that defaults occur independently from each other. In reality, it is known that this is not the case. The financial crisis of 2009 was a good example that defaults are correlated. The lack of correlation in the current prototypes could be resolved by fitting correlation between the defaults. Also correlation can be added by enlarging the scope of this research and adding macro-economic variables in the prototypes. The macro-economy is an important common factor for all the mortgage customers, and is expected to certify the main proportion of the correlation between the defaults. Adding macroeconomic variables will automatically make the prototypes more PIT. The impact of macroeconomic variables on the proportion of certified correlation could be an interesting topic for further research.

Another mismatch between the prototypes and reality is the interpretation of Life Time. As explained earlier, LT is in this research defined as a period of 5 years, because the restriction on available data for estimation and testing purposes. In reality LT could possibly be 30 years. The prototypes developed during this research are able to estimate the PD for a period of 30 years, because of the distributions assumed, but no statements could be made about the statistical significance of the performance for such a long period. Perhaps this long period performance could be tested by matching observation of customers who already are a long time in the portfolio with observations from younger customers. Developing and validating such an approach could be a subject for further research. Finding a successful approach for matching observations will also makes more data available for estimating LT prototypes.

The biggest gap between reality and the Logit prototype is probably the assumed Weibull distribution of defaults for LT estimations. It is not known if another distribution type will fit better, but this could be worthwhile trying. Another method that could improve the Logit prototype is to try different distributions, or combinations of distributions, for the different customer groups. In the current prototype is assumed that the different customer groups behave according to the same default distribution, but this does not have to match with reality. A deep dive into the behaviour of different customer groups could be a topic in further research.

The theoretical idea and the related assumptions made during this research to use the Tobit framework did not work out as intended, as result of a level problem caused by a bad fit. This problem is possibly caused by the high proportion of censored variables, low predictive power of variables, or a wrong assumption about a not continuously distributed depending variable. In the academically context it is interesting to stress the Tobit framework on the possible causes above in further research. For the practical application of the Tobit framework to predict PDs during this project it was not relevant and therefore not performed.

The promising Hazard prototype could very likely be improved by further research. In the current prototype a gap with reality exist caused by, among other things, the heuristic used for the selection of variables. This heuristic could be improved by further research. Besides that it could be valuable to investigate if the fit between the hazard rates and the realizations could be improved. Just as for the Logit LT prototype different distributions, or combinations of distributions, could be tried. Also another way of constructing the data set used, which result in more equally spread number of observations over time and customers groups and more observations in general, could be an interesting topic of a research.

Bibliography

A. Holst, D. W. (2014). *PARTICULIERE HYPOTHEKEN INTERNE RATING MODEL 2.0 (PHIRM 2.0) Datarapport.* Utrecht: SNS Bank N.V. .

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* , 716–723.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*(41), 997-1016.

B. A. Ciochetti, Y. D. (2002). The Termination of Mortgage Contracts through Prepayment and Default in the Commercial Mortgage Markets: A Proportional Hazard Approach with Competing Risks. *Real Estate Economics, 4*.

Bennett, S. (1993). Log-Logistic Regression Models for Survival Data. *Journal of the Royal Statistical Society*, 165-171.

Bretscher, O. (1995). *Linear Algebra With Applications.* Upper Saddle River: Prentice Hall.

C. Huber-Carol, N. B. (2002). *Goodness-of-Fit Tests and Model Validity.* Boston: Springer Science + Business Media.

C.B. Guure, N. I. (2002). ON PARTLY CENSORED DATA WITH THE WEIBULL DISTRIBUTION. *ARPN Journal of Engineering and Applied Sciences, 7*.

Cox, D. (1958). The Regression analysis of binary sequences. *J Roy Stat Soc B*, 215 - 242.

Dobson, A. (2002). *AN INTRODUCTION TO GENERALIZED LINEAR MODELS.* New York: Chapman & Hall/CRC.

G. Marsaglia, W. T. (2003). Evaluating Kolmogorov's Distribution. *Journal of Statistical Software*(18), 1-4.

Howell, D. (2007). *Statistical Methods for Psychology .* Belmont: Thomson Higher Education.

J. D. Kalbfleisch, R. L. (2002). *The Statistical Analysis of Failure Time Data.* Hoboken: Wiley & Sons.

KPMG. (2013). *New on the Horizon: Financial instruments - Expected dredit losses.* Retrieved from kpmg: kpmg.com/ifrs

Nederlandse bankensector. (n.d.). Retrieved 02 13, 2015, from Banken.nl: http://www.banken.nl/bankensector/bankensector-nederland

R. Chu, D. D. (2007). Best Practices for Managing Predictive Models in a Production Environment. *Data Mining and Predictive Modeling.* North Carolina: SAS Institute Inc.

Rodríguez, G. (2005). *Competing Risks.* Princeton: Princeton University.

Rodríguez, G. (2005). *Competing Risks.* Princeton: Princeton.edu.

Shumway, T. (1999). *Forecasting Bankruptcy More Accurately: A Simple Hazard Model.* Ann Arbor, Michigan, United States: University of Michigan Business School.

SNS Bank N.V. (2013). *COMPLIANCE MET AIRB TOEZICHTSEISEN VOOR KREDIETRISISCOMODELLEN - Intern document.* Utrecht: SNS Bank N.V.

SNS Bank N.V. (2013). *PARTICULIERE HYPOTHEKEN INTERNE RATING MODEL 2.0 (PHIRM 2.0).* Utrecht: SNS Bank N.V. .

SNS Bank N.V. (2014). *SNS Bank Annual report 2013.* Utrecht, the Netherlands.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*(240), 12851293.

Tasche, D. (2006). *Validation of internal rating systems and PD estimates.* Frankfurt am Main: Deutsche Bundesbank.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*(26), pp. 24-36.

Woolridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.

Wu, X. (2008). *Credit Scoring Model Validation.* Amsterdam: Universiteit van Amsterdam.

Zhang, D. (2005). *Likelihood and Censored (or Truncated) Survival Data.* Raleigh: North Carolina State University.

Appendices

APPENDIX 1 – List of Abbreviations

AUROC	- Area Under the Receiver Operating Characteristic curve
CIF	- Cumulative Incidence Function
DR	- Default Rate
CRIME	- Credit Risk Interactive Modelling Environment
FR&M	- Financial Risk & Modelling (department)
LT	- Life Time
LTFV	- Loan To Foreclosure Value
NHG	- Nationale Hypotheek Garantie
OOS	- Out of Sample
OOT	- Out of Time
PD	- Probability of Default
PHIRM 2.0 PD model	- Particuliere Hypotheken Interne Rating Probability of Default Model 2.0
PHIRM 3.0 PD model	- Particuliere Hypotheken Interne Rating Probability of Default Model 3.0
RWA	- Risk Weighted Assets

APPENDIX 2 – Regulations incorporated in this research

In this Appendix the regulations incorporated in this research are presented. These regulations are the same as used during the development of the PHIRM 2.0 PD model (SNS Bank N.V., 2013). Regulations that are not operative yet, but that will be incorporated in this research because it is likely that they will we active in the future are presented in APPENDIX 3. Because the data used in this research is given and bucketing is outside the scope of this research, all regulations regarding data and bucketing are neglected.

#	Description of regulations incorporated in this research
1.	The model has to contain borrower-, loan- and pledge characteristics, under the condition that that they have sufficient predictive power.
2.	The criteria for the classification of loans have to fit with the internal policy and the policy of the department “Arrears management”.
3.	Estimations of the PD have to be based on experience and empirical evidence and not on purely subjective considerations. The estimations also have to be intuitive and based on all relevant information. The less information is incorporated, the more conservative the estimates should be.
4.	Seasonal effects have to be observed and analysed.
5.	The model should allow for a meaningful differentiation of risk and provide accurate and consistent risk estimates.
6.	The model must have a good predictive power and no material deviation (bias).

APPENDIX 3 – Requirements and wishes given by the model owner and experts incorporated in this research

In this Appendix the requirements and wishes, as given by the model owner and the experts, who are incorporated in this research are presented. Regulations that are not operative yet, but that will be incorporated in this research because it is likely that they will be active in the future, are part of these requirements. Operative regulations incorporated in this research are presented in APPENDIX 2.

#	Description of requirements and wishes as given by the model owner and experts	Category
1.	For this research an alternative PD definition is developed that can be used for all the three frameworks. This makes the estimates from the different frameworks comparable. This definition is as follows: *“In fixed loans, there is a default in case if the total amount of arrears of a customer is equal to or greater than three times the monthly payable amount.*	Requirement
2.	The models that will be developed during this research have to be able to estimate Life Time (LT) PDs. LT is defined as: “*The period between the start and the end of a customer’s contract.”*	Requirement
3.	The difference between a LT PD and a 12-months PD has to be explainable.	Requirement
4.	The models that will be developed during this research must have a good predictive power, for instance an Area Under the Curve (AUC) of at least 80%.	Requirement
5.	The models don’t have to perform better on the quantitative aspect than the currently used PHIRM 2.0 PD model, under the condition that the difference in quantitative performance is compensated on the qualitative aspect.	Requirement
6.	The model’s predictive power has to be sufficient for different sections. These sections are: customers with Nationale Hypotheek Garantie (NHG) versus customers without NHG, customers from different geographical regions and customers from the different health groups used in the PHIRM 2.0 PD model (healthy, in arrears, default).	Requirement
7.	Used variables have to be selected based on their added value.	Requirement
8.	The variables that may be used have to be part of the short list or have to be used in the current PHIRM 2.0 PD model.	Requirement
9.	The model may be fitted on a sub portfolio, but the performance has to be comparable for the complete portfolio. The DBV portfolio is out of scope in this project, due to the fact that it is not comparable with the other portfolios.	Requirement
10.	The PD estimations have to be calculated at customer level and the model should enable the user to calculate them on a monthly basis.	Requirement
11.	The model has to represent the current situation, not the situation how it should be. Inefficient processes also have to be modelled.	Requirement
12.	All choices and assumptions made have to be documented sufficiently.	Requirement
13.	Used variables have to be intuitive.	Wish
14.	The used methods have to be explainable; this means that the department Credit Risk Retail can give a road show to demonstrate the model.	Wish

APPENDIX 4 – Dataset information

The dataset used is named 'ds_modellen_crpm.LOGITTOBITHAZARD_200709_201409' and is a combination of the tables A.Klant and B.Klant. The set contains, because of data quality issues, only observations from the period between August 2007 and October 2014. It is important to mention that the filters below are switched on during the development of the dataset.

```
WHERE       PHIRM_PZ_EXCLUDE_IND = 0
      AND   PHIRM_IRB_EXCLUDE_IND = 0
      AND   PHIRM_BLACKLIST_EXCLUDE_IND = 0
      AND   PHIRM_UITW_NAIJL_IND = 0
      AND   fraude_ind = 0
```

The explanation of most of the variables can be found in **PHIRM** 2.0 - **Data**rapport v1.01, available for insiders at the SNS Bank N.V. The explanation of the variables developed to be able to create the Tobit and Hazard prototype is given below (partly in pseudo code, SQL and Matlab code).

New Variable	Explanation
New_Default_score	`ACHTERSTAND_M_BG / (HOOFDSOM_M_BG/CONTRACTUELE_LOOPTIJD_MND_AANT))`
New_Default_Ind	IF New_Default_score <= -3 New_Default_ind = 1 ELSE New_Default_ind = 0 END
New_12_Months_Backward_num	This variable represents the number of New_Default_Ind defaults in the last 12 months, on the same way as the current defaults_ltst_12_mnd_aant variable.
New_12_Months_Forward_Ind	This variable represents the number of New_Default_Ind defaults in the upcoming 12 months, on the same way as the current in_default_komende_12mnd_ind variable.
New_24_Months_Forward_Ind	Same as New_12_Months_Forward_Ind but with 24 months instead of 12 months.
New_36_Months_Forward_Ind	Same as New_12_Months_Forward_Ind but with 36 months instead of 12 months.
New_60_Months_Forward_Ind	Same as New_12_Months_Forward_Ind but with 60 months instead of 12 months.
new_12_months_forward_highest_score	Lowest New_Default_score in upcoming 12 months.
new_24_months_forward_highest_score	Lowest New_Default_score in upcoming 24 months.
new_36_months_forward_highest_score	Lowest New_Default_score in upcoming 36 months.
new_60_months_forward_highest_score	Lowest New_Default_score in upcoming 60 months.
New_Healthy_Ind	`acht_ltst_12mnd_aant == 0 & New_12_Months_Backward_num == 0`
New_Recovered_Ind	`achterstand_ind == 0 & (acht_ltst_12mnd_aant > 0 | New_12_Months_Backward_num > 0)`
New_Arrears_Ind	`achterstand_ind == 1`
New_Logit_and_Tobit_Ind	WHERE PHIRM_PZ_EXCLUDE_IND = 0 AND PHIRM_IRB_EXCLUDE_IND = 0 AND PHIRM_BLACKLIST_EXCLUDE_IND = 0

	AND PHIRM_UITW_NAIJL_IND = 0 AND fraude_ind = 0
New_Hazard_Default_ind	`New_Hazard_Default_ind = New_Default_Ind`
New_Hazard_Prepayment_ind	Is equal to 1 if customer made a total prepayment, is zero if customer did not made a total prepeyment.
New_Hazard_Nothing_Observed_ind	Is equal to 1 if a customer did not made a total prepayment or did not went into default in his last observation, else zero.
New_Hazard_Combined_Ind	`New_Hazard_Combined_Ind = + New_Hazard_Nothing_Observed_ind + New_Hazard_Prepayment_ind + New_Hazard_Default_ind`
New_Hazard_Set_Ind	Each peil_dt where New_Hazard_Combined_Ind = 0 is removed as far as possible from New_Hazard_Combined_Ind = 1. See as example the table below and the explanation about the data selection for the Hazard prototype in paragraph 5.3.2.
New_Hazard_Censored_Default	If New_Hazard_Set_Ind = 1 and New_Hazard_Set_Ind is not equal to 1 as result of a default, than New_Hazard_Censored_Default = 1, else 0.
New_Hazard_Censored_PrePayment	If New_Hazard_Set_Ind = 1 and New_Hazard_Set_Ind is not equal to 1 as result of a prepayment, than New_Hazard_Censored_PrePayment = 1, else 0.
New_Hazard_Age_at_event_in_Months	looptijd_verstr_mnd_aant at peil_dt of event that results in New_Hazard_Set_Ind = 1 minus looptijd_verstr_mnd_aant at peil_dt which is removed as far as possible from the event peil_dt + 1.
New_Hazard_peil_dt_last_default	Peil_dt from last New_Hazard_Default_ind = 1 .
New_Hazard_num_defaults_ever	Number of New_Hazard_Default_ind = 1 ever observed.
New_Hazard_num_defaults_observ	Number of New_Hazard_Default_ind = 1 observed where peil_dt > 200709.

Customer	looptijd_verstr_mnd_aant	New_Hazard_Default_ind	New_Hazard_Prepayment_ind	New_Hazard_Nothing_Observed_ind	New_Hazard_Combined_Ind	New_Hazard_Set_Ind	New_Hazard_Censored_Default	New_Hazard_Censored_PrePayment	New_Hazard_Age_at_event_in_Months
1	1	0	0	0	0	1	1	1	6
	2	0	0	0	0	0	NaN	NaN	NaN
	3	0	0	0	0	0	NaN	NaN	NaN
	4	0	0	0	0	0	NaN	NaN	NaN
	5	0	0	0	0	0	NaN	NaN	NaN
	6	0	0	1	1	0	NaN	NaN	NaN
2	23	0	0	0	0	1	0	1	4
	24	0	0	0	0	0	NaN	NaN	NaN
	25	0	0	0	0	0	NaN	NaN	NaN
	26	1	0	0	1	0	NaN	NaN	NaN
	27	0	0	0	0	1	1	1	2
	28	0	0	1	1	0	NaN	NaN	NaN
3	6	1	0	0	1	0	NaN	NaN	NaN
	7	0	0	0	0	1	1	0	3
	8	0	0	0	0	0	NaN	NaN	NaN
	9	0	1	0	1	0	NaN	NaN	NaN

APPENDIX 5 – PHIRM PD 2.1 model bucket borders

This are the buckets as used in the PHIRM 2.0 PD model. The caps of the fifth, ninth and twelfth bucket are fixed manually. The higher border of bucket 12 is bigger than 1, because the model has to be able to cope with rounding errors from Matlab. The process of developing this bucket borders is described in the *PARTICULIERE HYPOTHEKEN INTERNE RATING MODEL 2.0* document.

Bucket	Lower border	Higher Border
1	0	0.0017068
2	0.0017068	0.0025186
3	0.0025186	0.0037766
4	0.0037766	0.0054915
5	0.0054915	0.0100000
6	0.0100000	0.0137780
7	0.0137780	0.0238170
8	0.0238170	0.0472700
9	0.0472700	0.1000000
10	0.1000000	0.1771100
11	0.1771100	0.3012800
12	0.3012800	1.0000001

APPENDIX 6 – Logit prototype variables

In this appendix the description of all selected variables for the Logit 2, 3 and 5 year prototypes are presented. The description of the 1 year Logit model variables can be found in paragraph 4.2.1. Also an overview of variables of the Tobit prototype with the variable names used inside SNS Bank N.V. is given.

1 year prototype (with SNS Bank N.V. variable names)

Sub Prototype, # of observations, Train/Test ratio, Scaling information	Characteristic type	Variable name	Value
Healthy, n = 581388, (70/30), X variable normalized	Borrower	achterstanden_totaal_aant	$\beta_1 = +0.20157$
		defaults_ltst_36mnd_aant	$\beta_2 = +0.0936576$
	Loan	gemiddeld_rente_pct	$\beta_3 = +0.27011$
	Pledge	saldo_erh_ck_bg	$\beta_4 = -0.122986$
		hoofdsom_nhg_m_bg	$\beta_5 = -0.217029$
		ltfv_m_vb_100_pct	$\beta_6 = +0.499391$
	Constant	-	$\beta_0 = -5.7205$
Recovered, n=32422, (70/30), X variable normalized	Borrower	acht_ltst_6mnd_aant	$\beta_1 = +0.334063$
		achterstanden_totaal_aant	$\beta_2 = +0.223352$
		defaults_ltst_36mnd_aant	$\beta_3 = +0.327995$
	Loan	gemiddeld_rente_pct	$\beta_4 = +0.277961$
	Constant	-	$\beta_0 = -2.6982$
Arrears, n= 13184, (70/30), X variable normalized	Borrower	defaults_ltst_36mnd_aant	$\beta_1 = +0.385075$
		new_default_score	$\beta_2 = -0.689566$
	Loan	betreg_m_ind	$\beta_3 = -0.319641$
	Pledge	ltfv_m_vb_100_pct	$\beta_4 = +0.153832$
	Constant	-	$\beta_0 = -1.0931$

2 year prototype (with SNS Bank N.V. variable names)

Sub Prototype, # of observations, Test/Train ratio, Scaling information	Characteristic type	Variable name	Value
Healthy, n = 466059 (70/30) X var scaled	Borrower	achterstanden_totaal_aant	$\beta_1 = +0.236889$
	Loan	gemiddeld_rente_pct	$\beta_2 = +0.190762$
	Pledge	hoofdsom_nhg_m_bg	$\beta_3 = -0.209293$
		ltfv_m_vb_100_pct	$\beta_4 = +0.581422$
	Constant	-	$\beta_0 = -4.8172$
Recovered, n= 22702 (70/50) X var scaled	Borrower	acht_ltst_6mnd_aant	$\beta_1 = +0.301322$
		achterstanden_totaal_aant	$\beta_2 = +0.254836$
		defaults_ltst_36mnd_aant	$\beta_3 = +0.315902$
	Loan	gemiddeld_rente_pct	$\beta_4 = +0.204489$
	Pledge	ltfv_m_vb_100_pct	$\beta_5 = +0.194525$
	Constant	-	$\beta_0 = -1.9698$
Arrears, n= 9190 (70/50) X var scaled	Borrower	defaults_ltst_36mnd_aant	$\beta_1 = +0.435905$
		new_default_score	$\beta_2 = -0.604254$
	Loan	betreg_m_ind	$\beta_3 = -0.261611$
	Pledge	ltfv_m_vb_100_pct	$\beta_4 = +0.178476$
	Constant	-	$\beta_0 = -0.58727$

2 year prototype

Sub Prototype, # of observations, Test/Train ratio, Scaling information	Character -istic type	Variable description	Value
Healthy, n = 466059 (70/30) X var scaled	Borrower	Total number of arrears (#)	β_1 = +0.236889
	Loan	Average interest (%)	β_2 = +0.190762
	Pledge	NHG principal (€)	β_3 = -0.209293
		Loan to foreclosure value (%)	β_4 = +0.581422
	Constant	-	β_0 = -4.8172
Recovered, n= 22702 (70/50) X var scaled	Borrower	Number of arrears in the last 6 months (#)	β_1 = +0.301322
		Total number of arrears (#)	β_2 = +0.254836
		Number of defaults in the last 36 months (#)	β_3 = +0.315902
	Loan	Average interest (%)	β_4 = +0.204489
	Pledge	Loan to foreclosure value (%)	β_5 = +0.194525
	Constant	-	β_0 = -1.9698
Arrears, n= 9190 (70/50) X var scaled	Borrower	Number of defaults in the last 36 months (#)	β_1 = +0.435905
		Amount arrears / month term (€)	β_2 = -0.604254
	Loan	Payment arrangement indicator {0,1}	β_3 = -0.261611
	Pledge	Loan to foreclosure value (%)	β_4 = +0.178476
	Constant	-	β_0 = -0.58727

3 year prototype (with SNS Bank N.V. variable names)

Sub Prototype, # of observations, Test/Train ratio, Scaling information	Characteristic type	Variable name	Value
Healthy, n= 348647 (70/30) X var scaled	Borrower	achterstanden_totaal_aant	β_1 = +0.246484
	Loan	gemiddeld_rente_pct	β_2 = +0.1773
	Pledge	hoofdsom_nhg_m_bg	β_3 = -0.234259
		ltfv_m_vb_100_pct	β_4 = +0.631365
		saldo_erh_ck_bg	β_5 = -0.109682
	Constant	-	β_0 = -4.3623
Recovered, n = 20084 (70/30) X vars scaled	Borrower	acht_ltst_6mnd_aant	β_1 = +0.255062
		achterstanden_totaal_aant	β_2 = +0.285632
		defaults_ltst_36mnd_aant	β_3 = +0.331241
	Pledge	ltfv_m_vb_100_pct	β_4 = +0.198364
	Constant	-	β_0 = -1.6097
Arrears, n = 7309 (70/30) X vars scaled	Borrower	acht_ltst_12mnd_aant	β_1 = +0.266204
		defaults_ltst_36mnd_aant	β_2 = +0.387315
		new_default_score	β_3 = -0.506723
	Loan	bb_lening_ind	β_4 = +0.178788
		betreg_m_ind	β_5 = -0.150512
	Pledge	ltfv_m_vb_100_pct	β_6 = +0.221005
	Constant	-	β_0 = -0.3902

3 year prototype

Sub Prototype, # of observations, Test/Train ratio, Scaling information	Character -istic type	Variable description	Value
Healthy, n= 348647 (70/30) X var scaled	Borrower	Total number of arrears (#)	β_1 = +0.246484
	Loan	Average interest (%)	β_2 = +0.1773
	Pledge	NHG principal (€)	β_3 = -0.234259
		Loan to foreclosure value (%)	β_4 = +0.631365
		Consumer Credit part of extra mortgage (€)	β_5 = -0.109682
	Constant	-	β_0 = -4.3623
Recovered, n = 20084 (70/30) X vars scaled	Borrower	Number of arrears in the last 6 months (#)	β_1 = +0.255062
		Total number of arrears (#)	β_2 = +0.285632
		Number of defaults in the last 36 months (#)	β_3 = +0.331241
	Pledge	Loan to foreclosure value (%)	β_4 = +0.198364
	Constant	-	β_0 = -1.6097
Arrears, n = 7309 (70/30) X vars scaled	Borrower	Number of arrears in the last 12 months (#)	β_1 = +0.266204
		Number of defaults in the last 36 months (#)	β_2 = +0.387315
		Amount arrears / month term (€)	β_3 = -0.506723
	Loan	Arrears management loan indicator {0,1}	β_4 = +0.178788
		Payment arrangement indicator {0,1}	β_5 = -0.150512
	Pledge	Loan to foreclosure value (%)	β_6 = +0.221005
	Constant	-	β_0 = -0.3902

5 year prototype (with SNS Bank N.V. variable names)

Sub Prototype, # of observations, Test/Train ratio, Scaling information	Characteristic type	Variable description	Value
Healthy, n = 115077 (70/50) X vars scaled	Borrower	achterstanden_totaal_aant	β_1 = +0.244678
	Loan	intermediair_ind	β_2 = +0.173446
	Pledge	hoofdsom_nhg_m_bg	β_3 = -0.275103
		ltfv_m_vb_100_pct	β_4 = +0.590617
	Constant	-	β_0 = -3.8107
Recovered, n = 6817 (50/50) X vars scaled	Borrower	acht_ltst_6mnd_aant	β_1 = +0.243091
		achterstanden_totaal_aant	β_2 = +0.352652
		defaults_ltst_36mnd_aant	β_3 = +0.293974
	Pledge	ltfv_m_vb_100_pct	β_4 = +0.315859
	Constant	-	β_0 = -1.2651
Arrears, n= 2332 (70/30) X vars scaled	Borrower	acht_ltst_12mnd_aant	β_1 = +0.287139
		defaults_ltst_36mnd_aant	β_2 = +0.406856
		new_default_score	β_3 = -0.56254
	Loan	betreg_m_ind	β_4 = -0.0597819
	Pledge	ltfv_m_vb_100_pct	β_5 = +0.169788
	Constant	-	β_0 = -0.11293

5 year prototype

Sub Prototype, # of observations, Test/Train ratio, Scaling information	Character -istic type	Variable name	Value
Healthy, n = 115077 (70/50) X vars scaled	Borrower	Total number of arrears (#)	β_1 = +0.244678
	Loan	Intermediary indicator {0,1}	β_2 = +0.173446
	Pledge	NHG principal (€)	β_3 = -0.275103
		Loan to foreclosure value (%)	β_4 = +0.590617
	Constant	-	β_0 = -3.8107
Recovered, n = 6817 (50/50) X vars scaled	Borrower	Number of arrears in the last 6 months (#)	β_1 = +0.243091
		Total number of arrears (#)	β_2 = +0.352652
		Number of defaults in the last 36 months (#)	β_3 = +0.293974
	Pledge	Loan to foreclosure value (%)	β_4 = +0.315859
	Constant	-	β_0 = -1.2651
Arrears, n= 2332 (70/30) X vars scaled	Borrower	Number of arrears in the last 12 months (#)	β_1 = +0.287139
		Number of defaults in the last 36 months (#)	β_2 = +0.406856
		Amount arrears / month term (€)	β_3 = -0.56254
	Loan	Payment arrangement indicator {0,1}	β_4 = -0.0597819
	Pledge	Loan to foreclosure value (%)	β_5 = +0.169788
	Constant	-	β_0 = -0.11293

APPENDIX 7 – Tobit prototype variables

This APPENDIX contains the overview of variables of the Tobit prototype with the variable names used inside SNS Bank N.V.

Sub Prototype, # of observations, Train/Test ratio, Scaling information	Characteristic type	Variable name	Value
Healthy, n = 396176, (50/50), X variable Scaled	Borrower	achterstanden_totaal_aant	β_1 = +0.711965
	Loan	looptijd_verstr_oudst_mnd_aant	β_2 = -0.657999
		Hypotheekgarantie_ind	β_3 = -0.282072
	Pledge	ltfv_m_vb_100_pct	β_4 = +0.331212
	Constant	-	β_0 = -7.5003
	Sigma	-	σ = 3.9756
Recovered, n = 21898, (50/50), X variable Scaled	Borrower	acht_ltst_6mnd_aant	β_1 = +0.493453
		achterstanden_totaal_aant	β_2 = +0.508115
		defaults_ltst_36mnd_aant	β_3 = +0.265358
		gem_duur_acht_12mnd_aant	β_4 = +0.130338
	Loan	bb_lening_ind	β_5 = +0.131762
	Pledge	ltfv_m_vb_100_pct	β_6 = +0.164622
	Constant	-	β_0 = -0.22569
	Sigma	-	σ = 2.4920
Arrears, n = 8963, (50/50), X variable Scaled	Borrower	defaults_ltst_36mnd_aant	β_1 = +0.506033
		new_default_score	β_2 = -0.764544
	Loan	bb_lening_ind	β_3 = +0.302402
	Pledge	ltfv_m_vb_100_pct	β_4 = +0.130984
	Constant	-	β_0 = 2.2212
	Sigma	-	σ = 2.4124

APPENDIX 8 – Hazard prototype variables

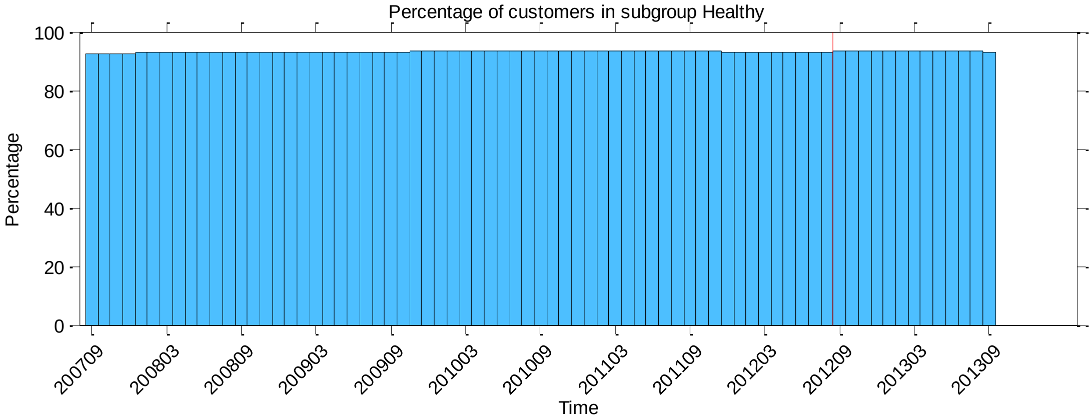
This APPENDIX contains the overview of variables of the Hazard prototype with the variable names used inside SNS Bank N.V.

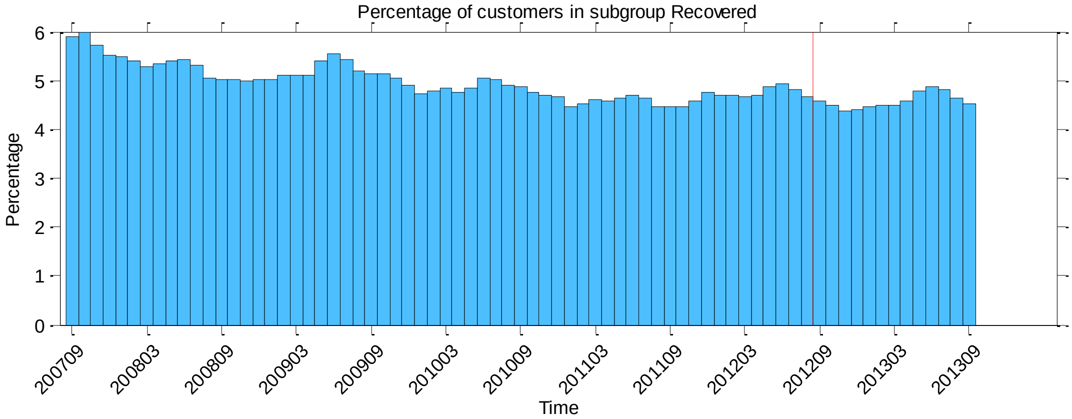
Sub Prototype, # of observations, Scaling information	Event	Characteristic type	Variable name	Value
Healthy, n = 226290, X variables scaled on [0,1]	Default	Borrower	new_hazard_num_defaults_ever	β_1 = -11.551969
			achterstanden_totaal_aant	β_2 = -4.220677
			defaults_ltst_36mnd_aant	β_3 = 1.777056
		Loan	gemiddeld_rente_pct	β_4 = -1.193231
		Pledge	saldo_erh_ck_bg	β_5 = -0.070829
			hoofdsom_nhg_m_bg	β_6 = 0.586302
			ltfv_m_vb_100_pct	β_7 = -3.342875
		Constant	-	β_0 = 9.217739
		Alpha	-	α_{def} = 0.761171
	Prepayment	Tau	-	τ_{prep} = 209.641623
		Alpha	-	α_{prep} = 1.340222
Recovered, n = 17817, X variables scaled on [0,1]	Default	Borrower	new_hazard_num_defaults_ever	β_1 = -6.645824
			achter_3mnd_ind	β_2 = -0.399239
			achterstanden_totaal_aant	β_3 = -1.924554
			defaults_ltst_36mnd_aant	β_4 = -0.260921
			acht_ltst_6mnd_aant	β_5 = -0.864793
		Loan	gemiddeld_rente_pct	β_6 = -1.74342
		Pledge	ltfv_m_vb_100_pct	β_7 = -1.139098
		Constant	-	β_0 = 7.245091
		Alpha	-	α_{def} = 0.835663
	Prepayment	Tau	-	τ_{prep} = 221.297099
		Alpha	-	α_{prep} = 1.097646
Arrears, n = 5587, X variables scaled on [0,1]	Default	Borrower	new_hazard_num_defaults_ever	β_1 = -5.283660
			defaults_ltst_36mnd_aant	β_2 = -0.058718
			new_default_score	β_3 = 1.291860
			acht_ltst_12mnd_aant	β_4 = -0.945830
		Loan	bb_lening_ind	β_5 = -0.524896
			betreg_m_ind	β_6 = 0.1866110
		Pledge	ltfv_m_vb_100_pct	β_7 = -1.253996
		Constant	-	β_0 = 4.406391
		Alpha	-	α_{def} = 0.708987
	Prepayment	Tau	-	τ_{prep} = 208.118199
		Alpha	-	α_{prep} = 1.2442940

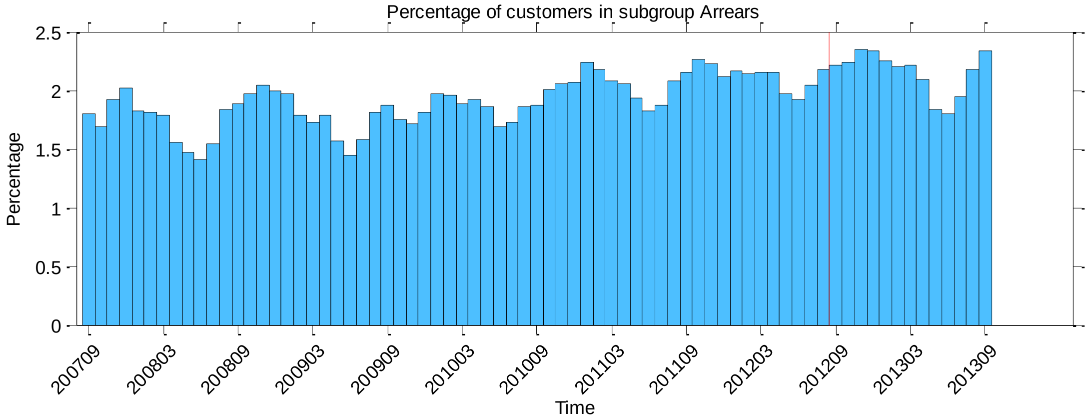
APPENDIX 9 – Logit 1 year prototype results

This APPENDIX contains statistics and test results from the Logit 1 year prototype. The mean number of observations per moment of measurement is equal to 161050.4247.

Healthy/Recovered/Default

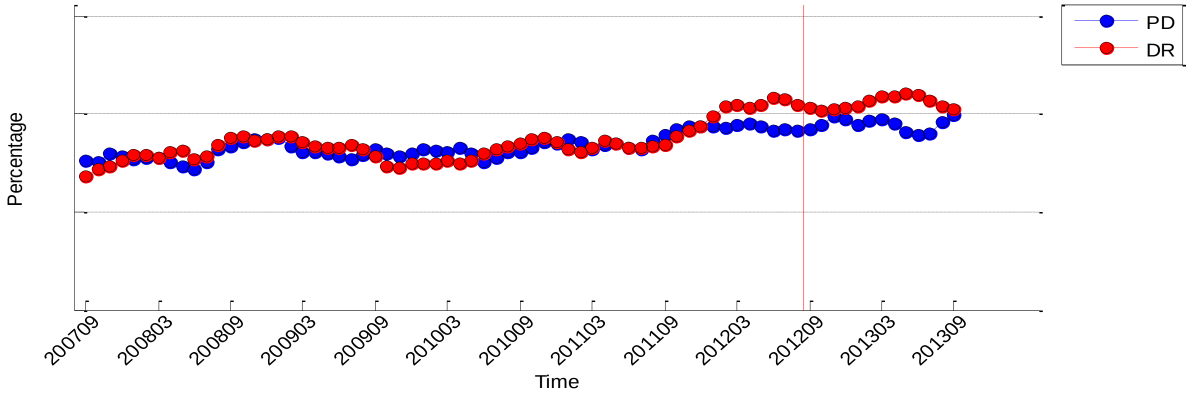




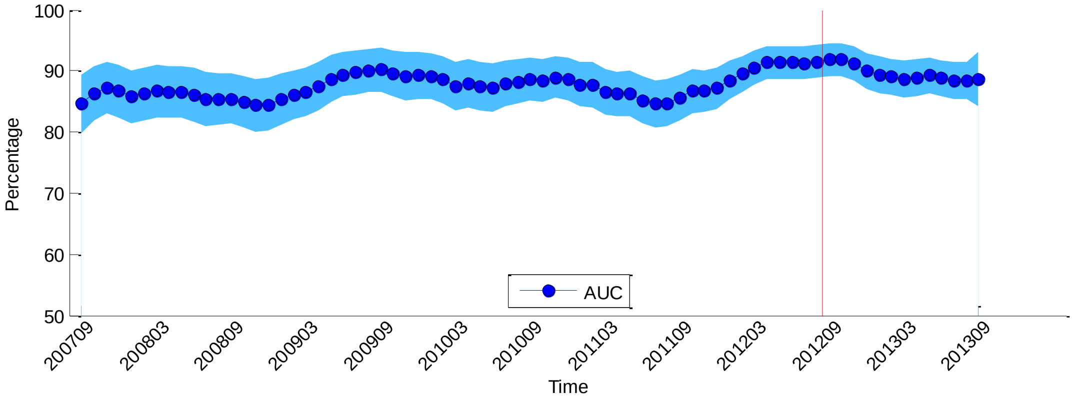


NHG/Non NHG

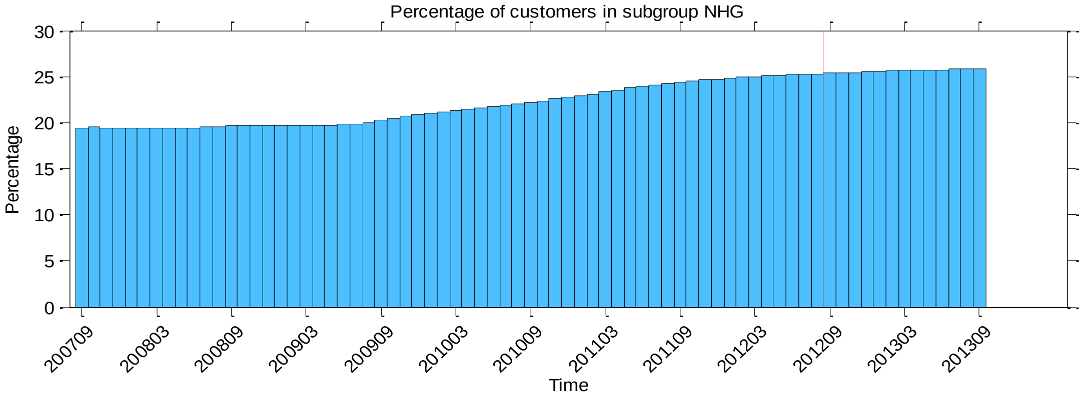
PD and DR through time - NHG



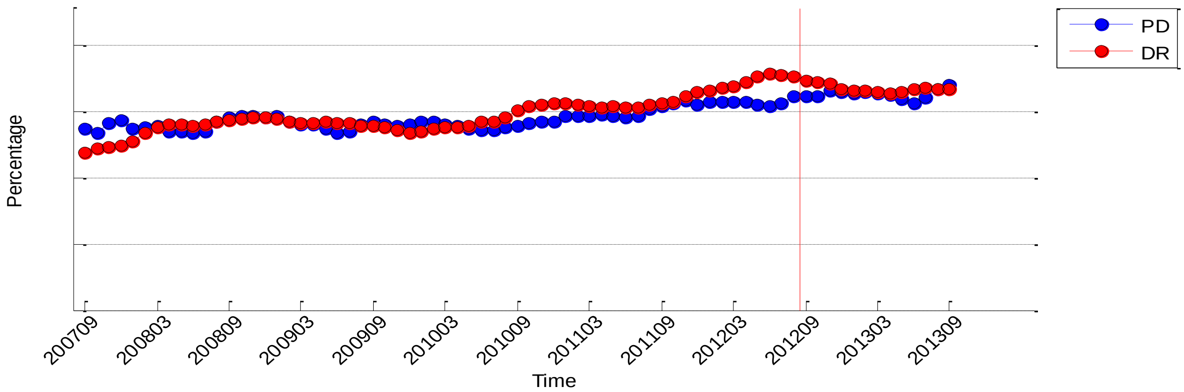
AUC through time - NHG

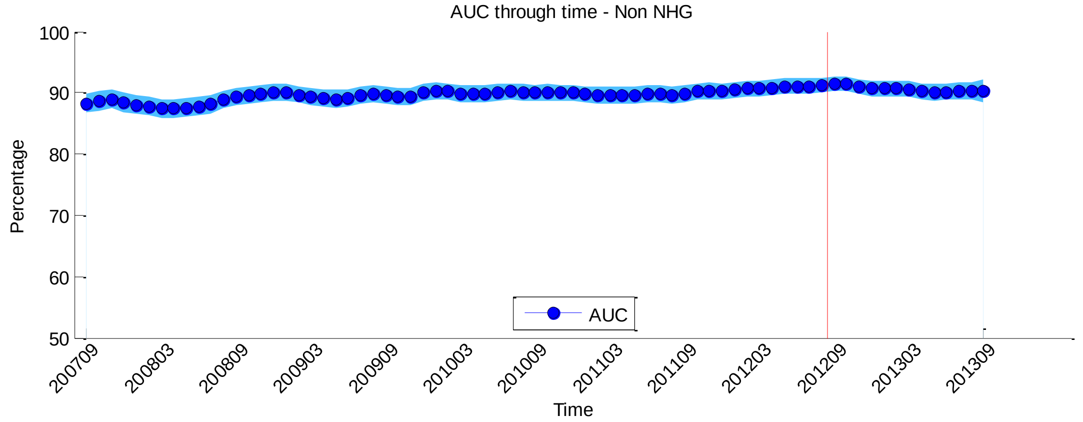


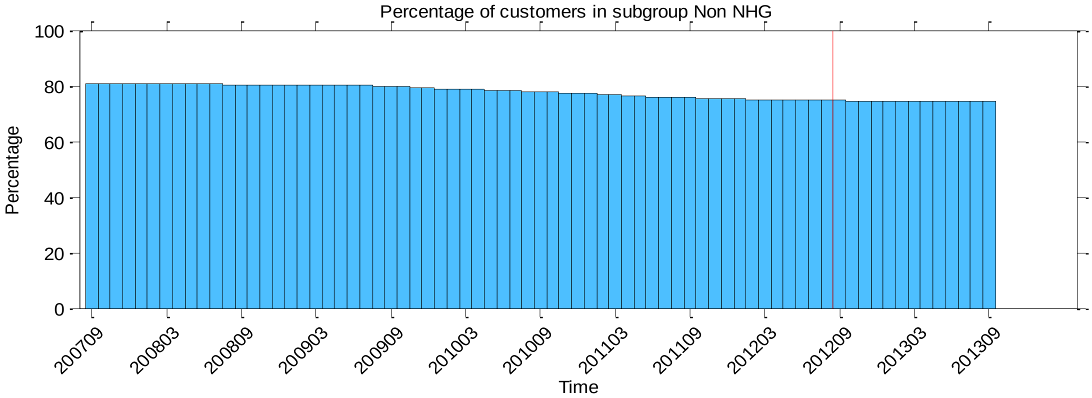
Percentage of customers in subgroup NHG



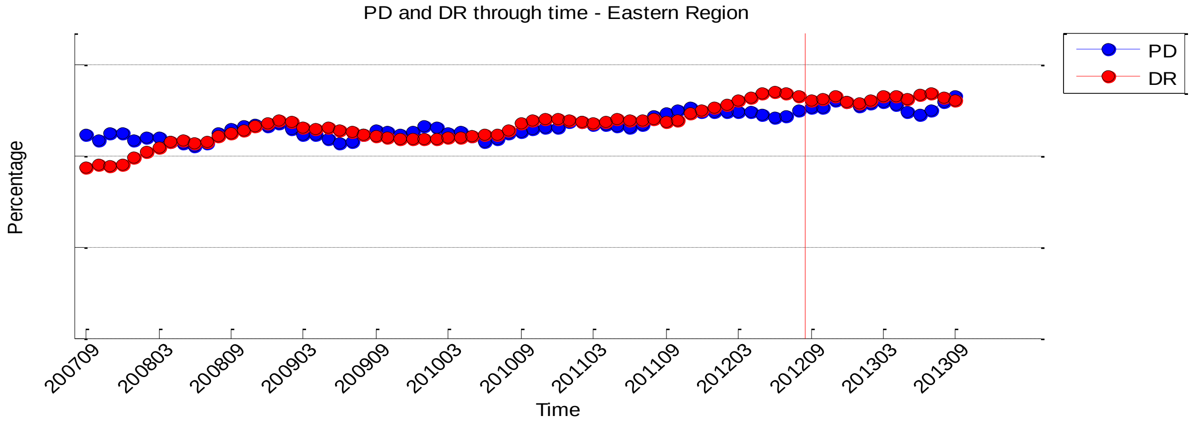
PD and DR through time - Non NHG

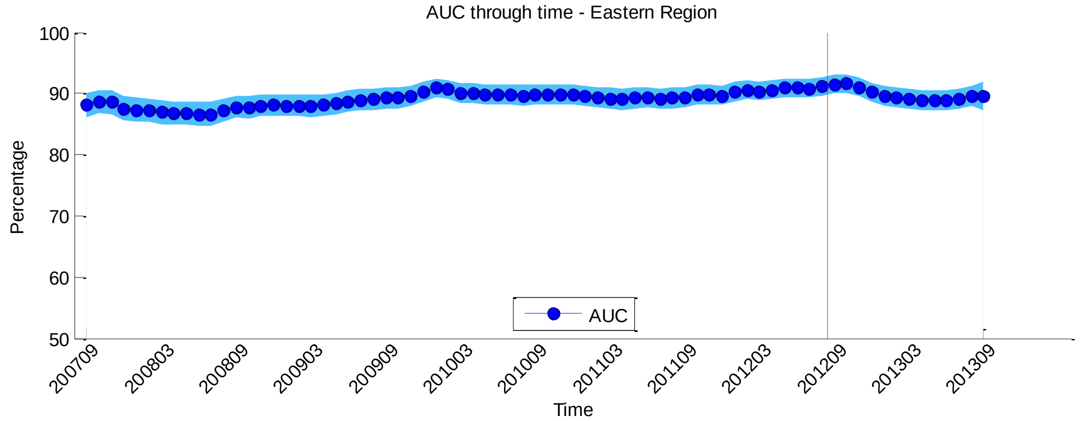


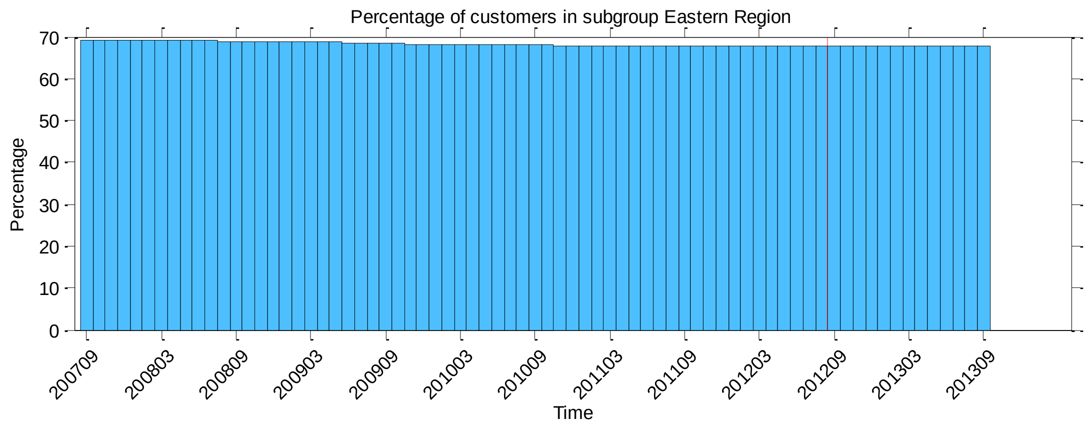

AUC through time - Non NHG
Percentage
100
90
80
70
60
50
AUC
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309
Time

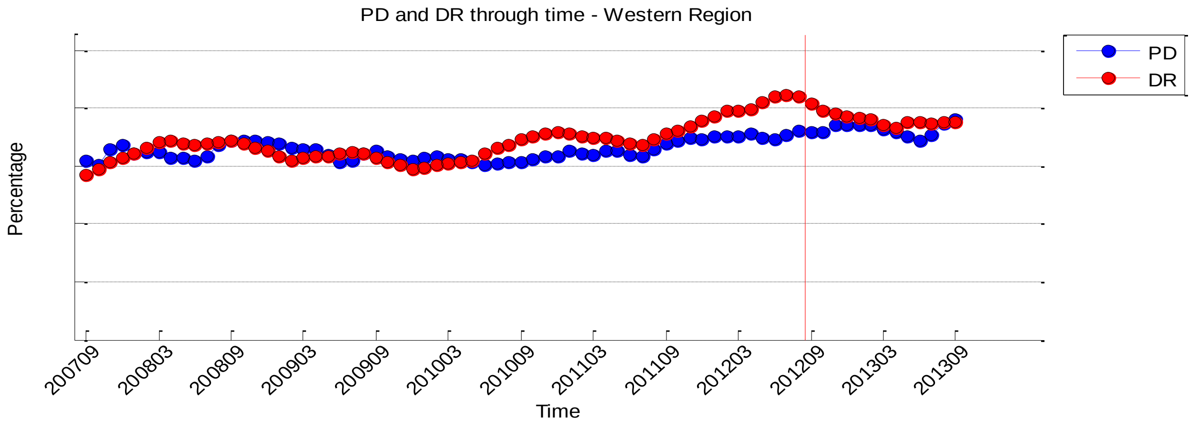

Percentage of customers in subgroup Non NHG
Percentage
100
80
60
40
20
0
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309
Time

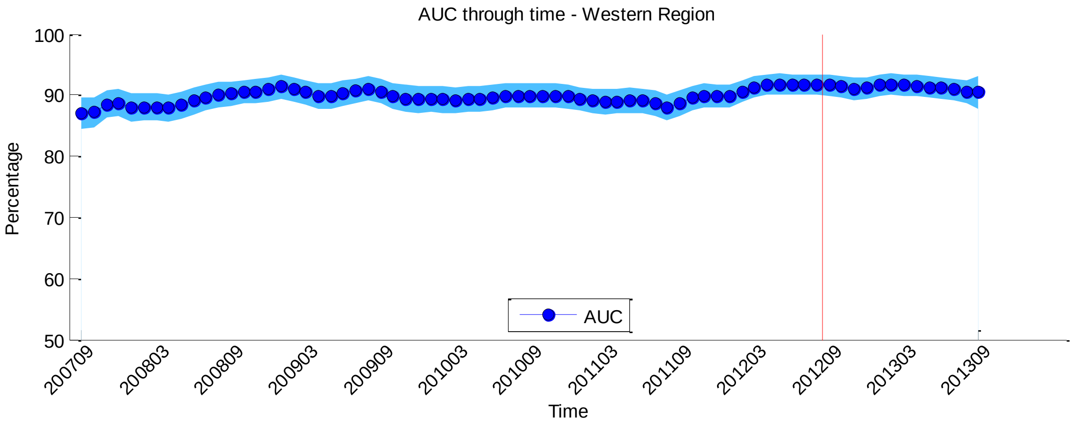
Eastern Region/Western Region

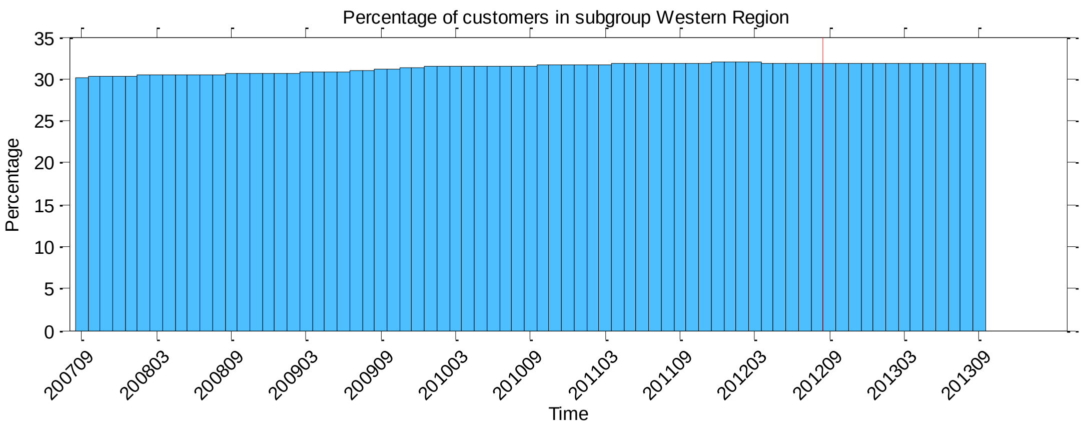




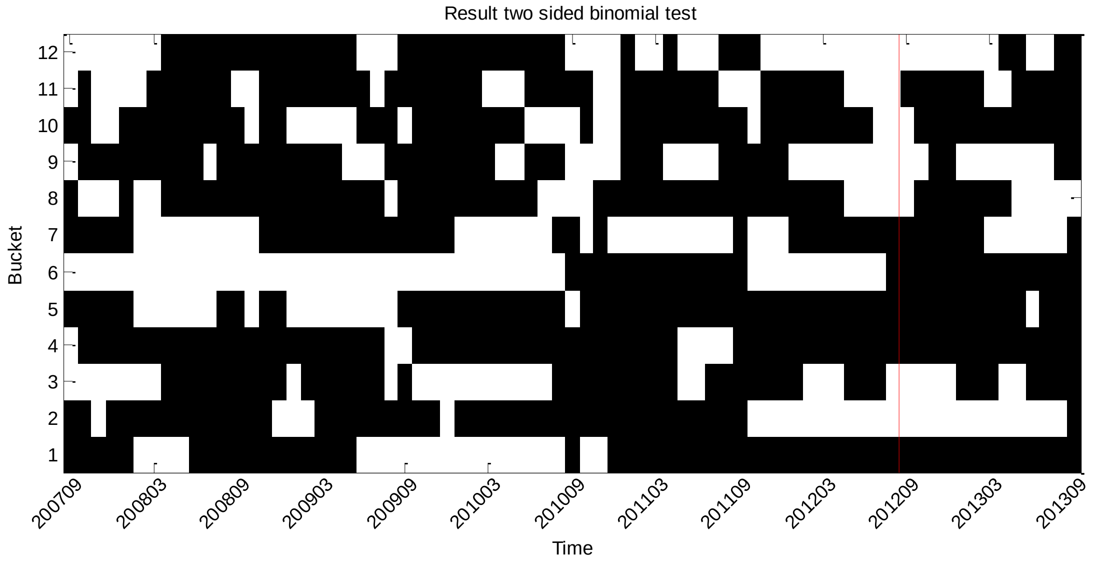



PD and DR through time - Western Region
PD
DR
Percentage
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309
Time

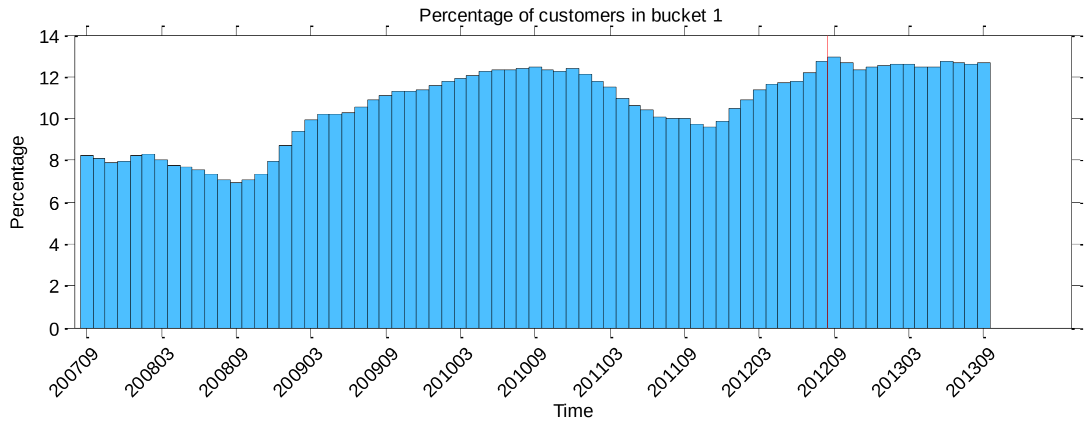

AUC through time - Western Region
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309
Time

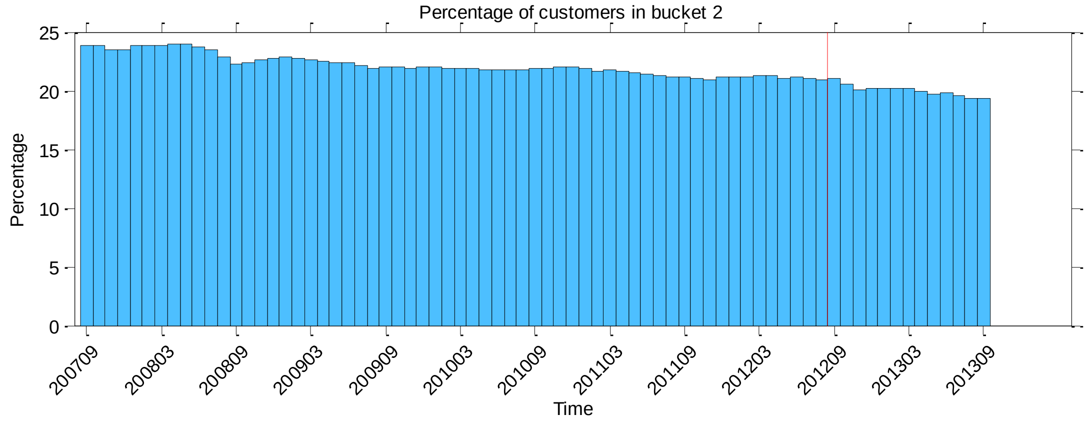

Percentage of customers in subgroup Western Region
35
30
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309
Time

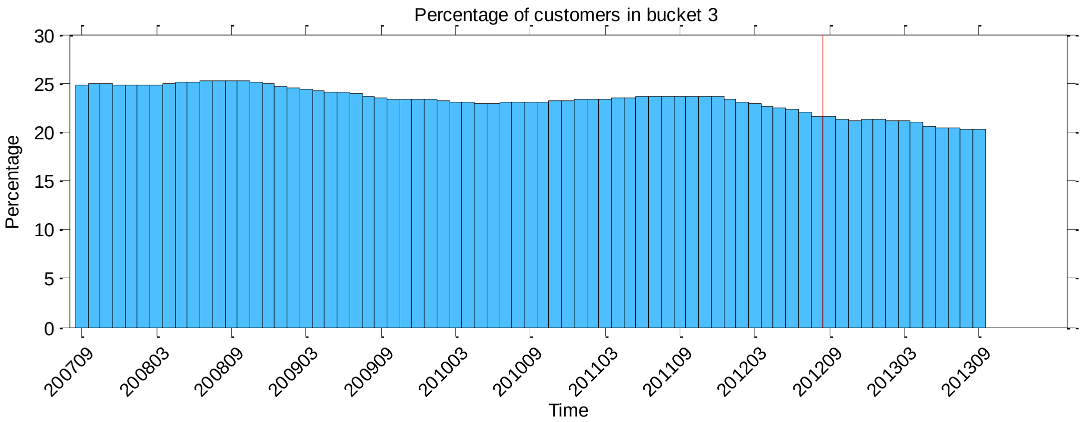
Binomial test Results on bucket level

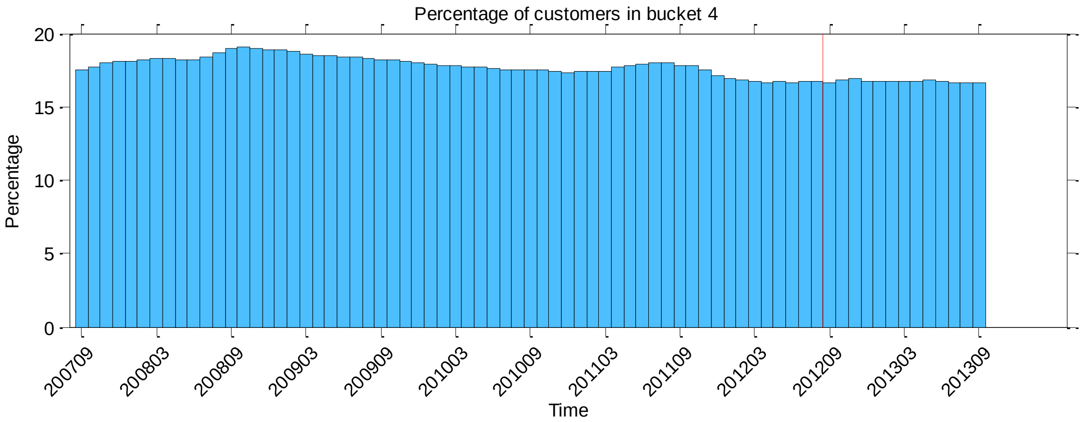


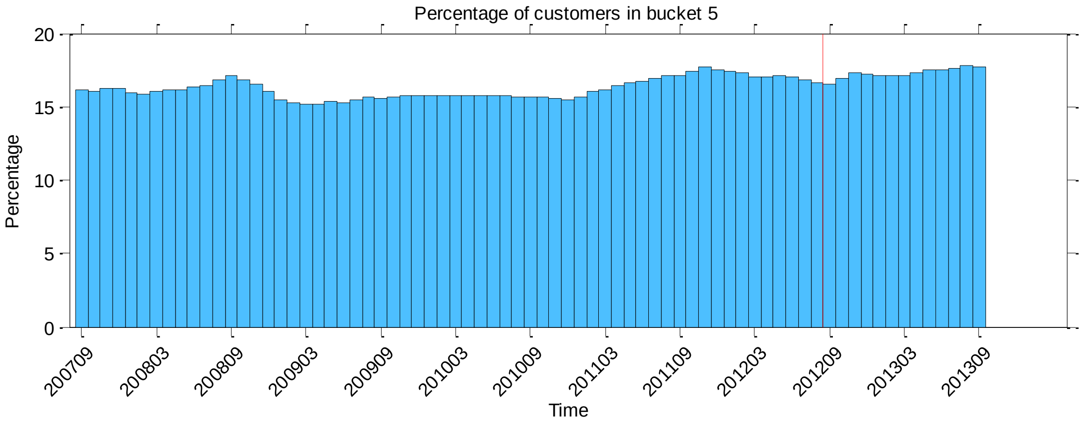
Percentage of customers per bucket

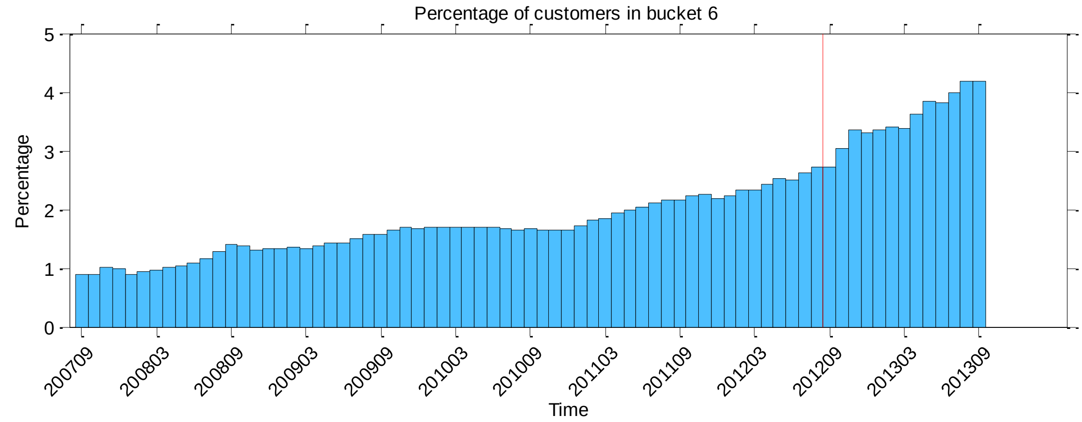


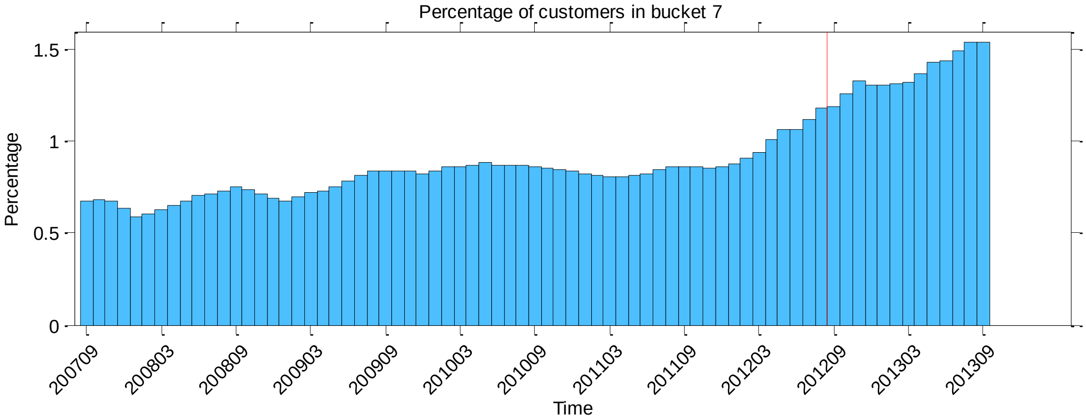


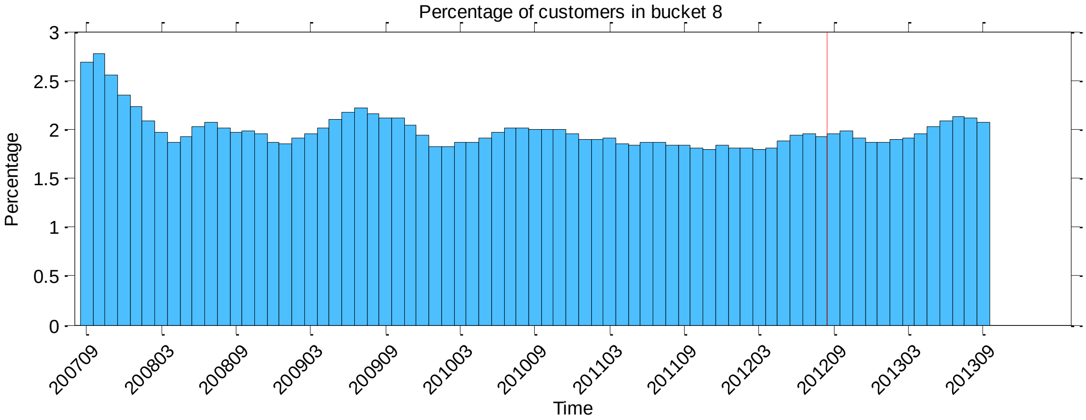

Percentage of customers in bucket 3
Percentage
Time
200709 200803 200809 200903 200909 201003 201009 201103 201109 201203 201209 201303 201309

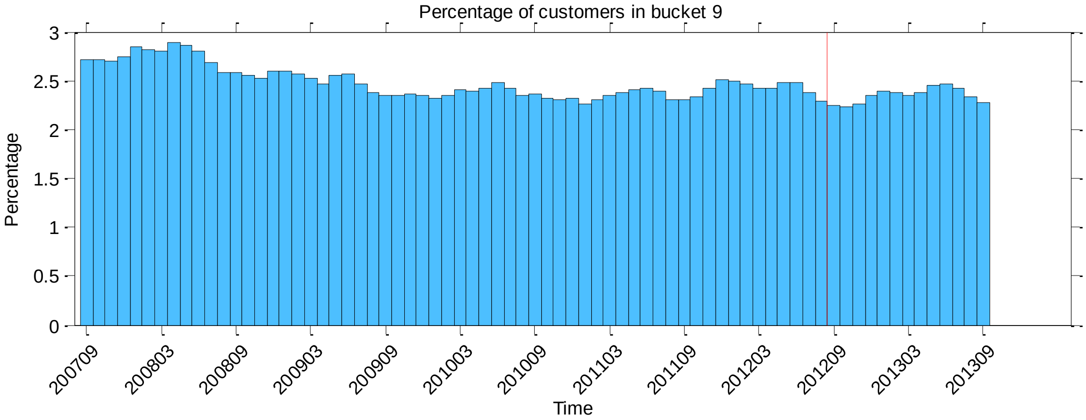

Percentage of customers in bucket 4
Percentage
Time
200709 200803 200809 200903 200909 201003 201009 201103 201109 201203 201209 201303 201309

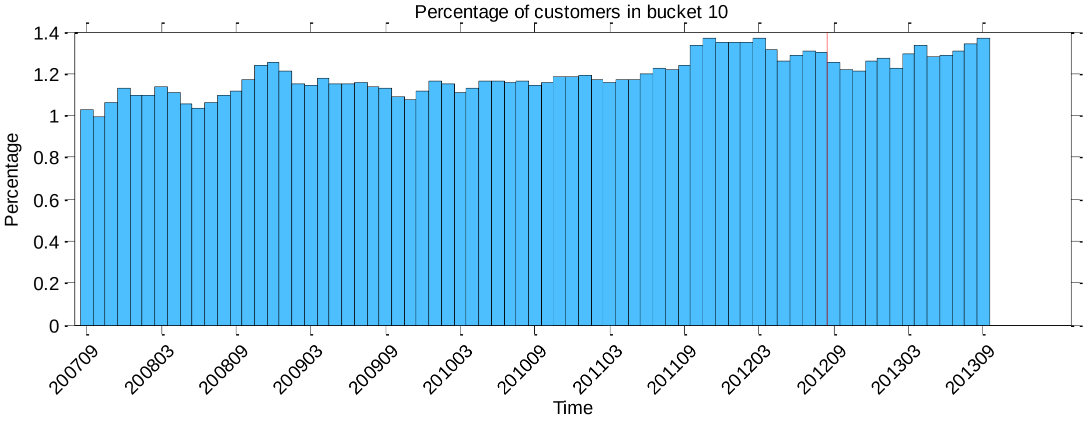

Percentage of customers in bucket 5
Percentage
Time
200709 200803 200809 200903 200909 201003 201009 201103 201109 201203 201209 201303 201309

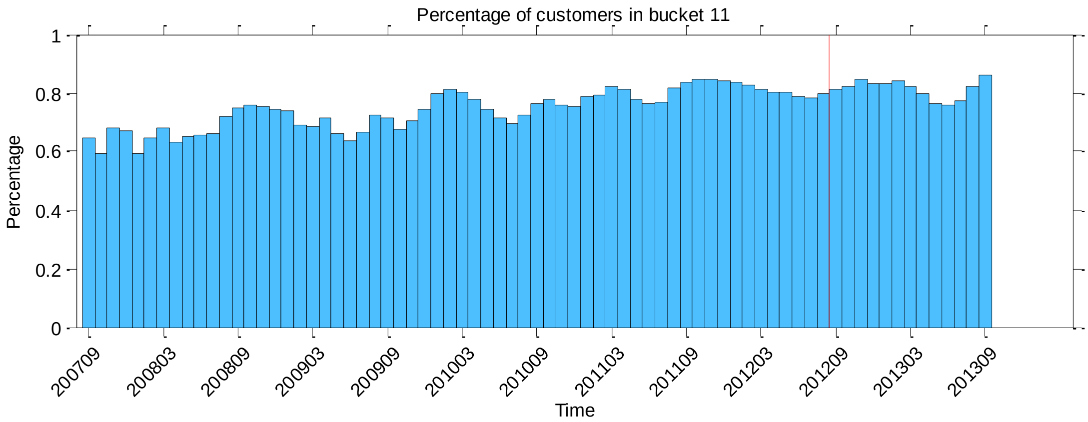

Percentage of customers in bucket 6
Percentage
Time
200709 200803 200809 200903 200909 201003 201009 201103 201109 201203 201209 201303 201309

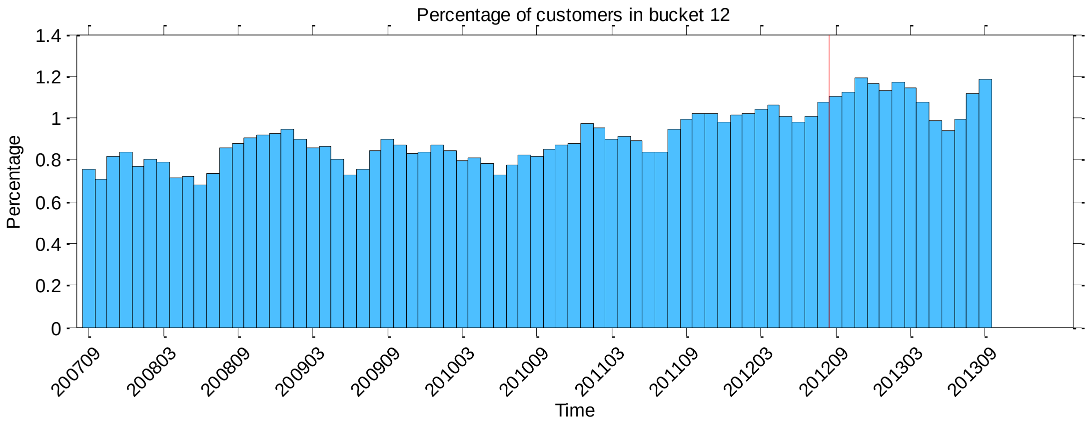

Percentage of customers in bucket 7
Percentage
Time
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309


Percentage of customers in bucket 8
Percentage
Time
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309


Percentage of customers in bucket 9
Percentage
Time
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309


Percentage of customers in bucket 10
Percentage
Time
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309

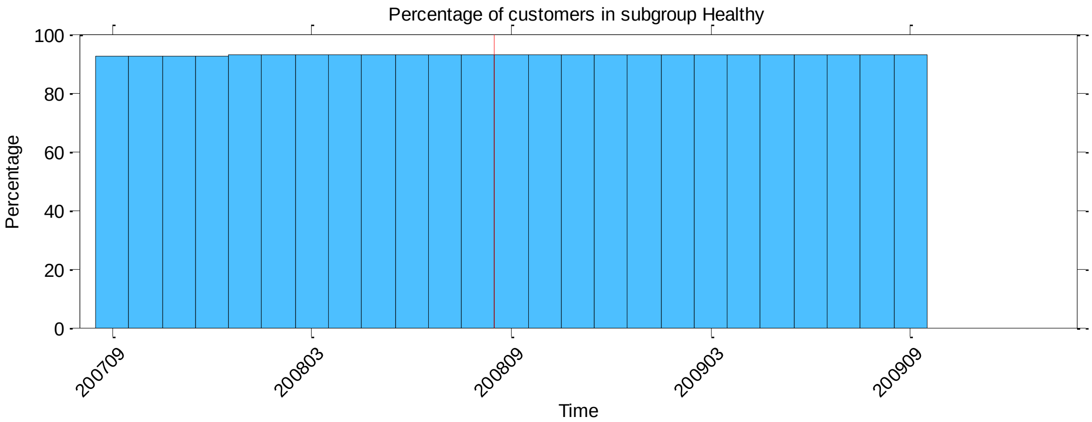

Percentage of customers in bucket 11
Percentage
1
0.8
0.6
0.4
0.2
0
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309
Time

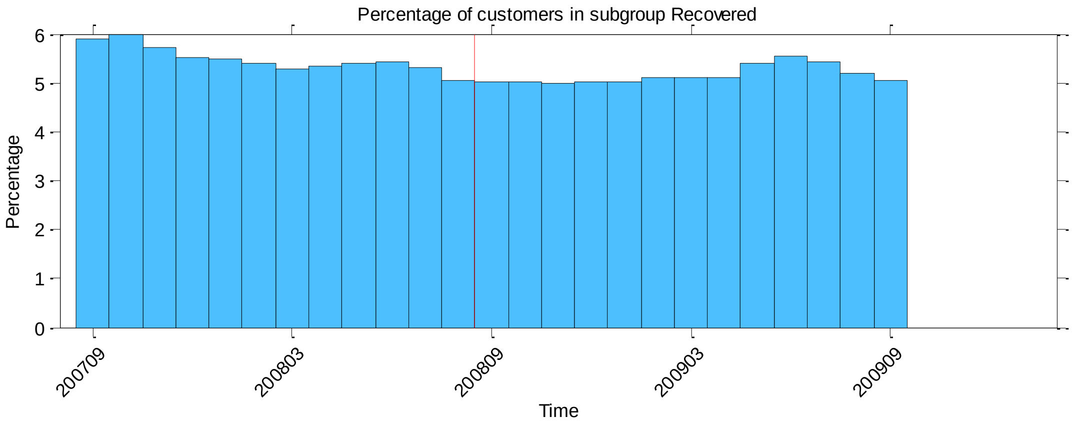

Percentage of customers in bucket 12
Percentage
1.4
1.2
1
0.8
0.6
0.4
0.2
0
200709
200803
200809
200903
200909
201003
201009
201103
201109
201203
201209
201303
201309
Time

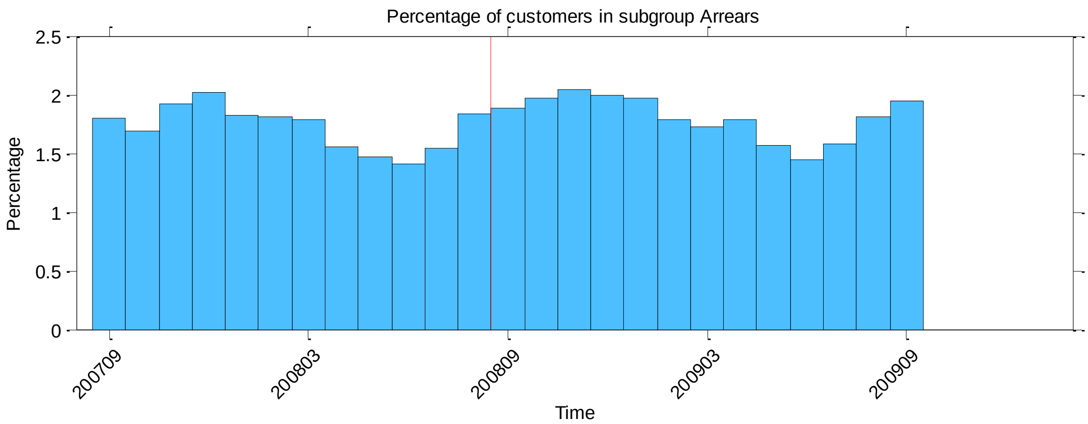
APPENDIX 10 – Logit LT prototype with mean Lambda extrapolation results

This APPENDIX contains statistics and test results from the Logit LT prototype with mean Lambda extrapolation. The mean number of observations per moment of measurement is equal to 159381.56.

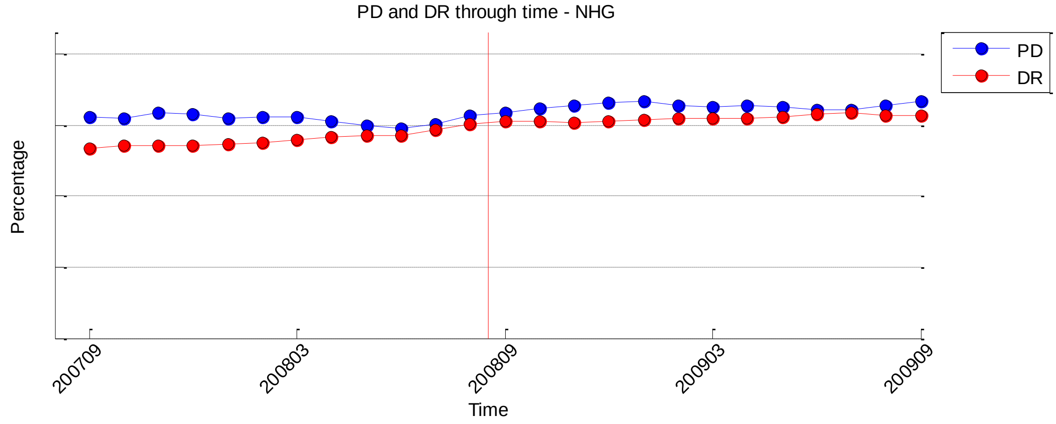
Healthy/Recovered/Default

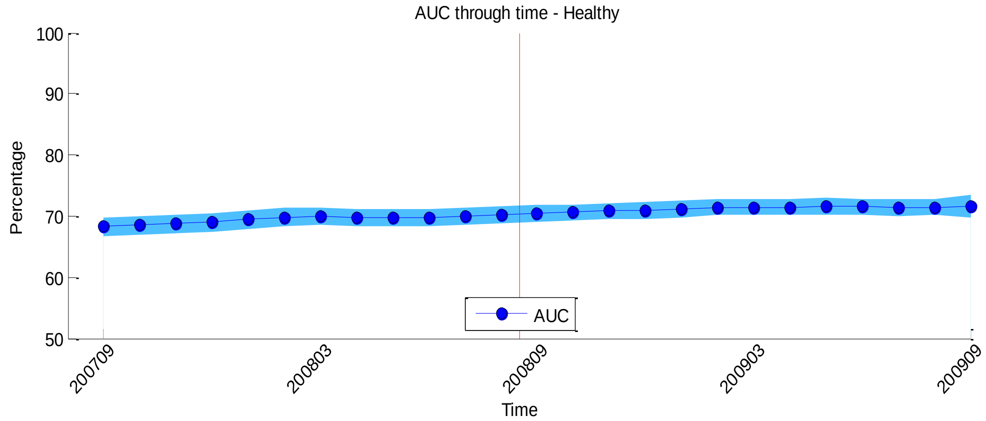


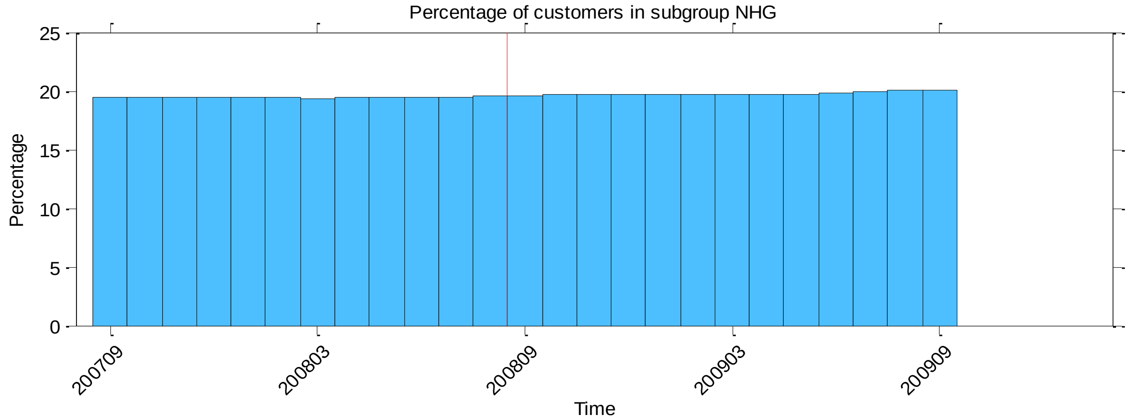


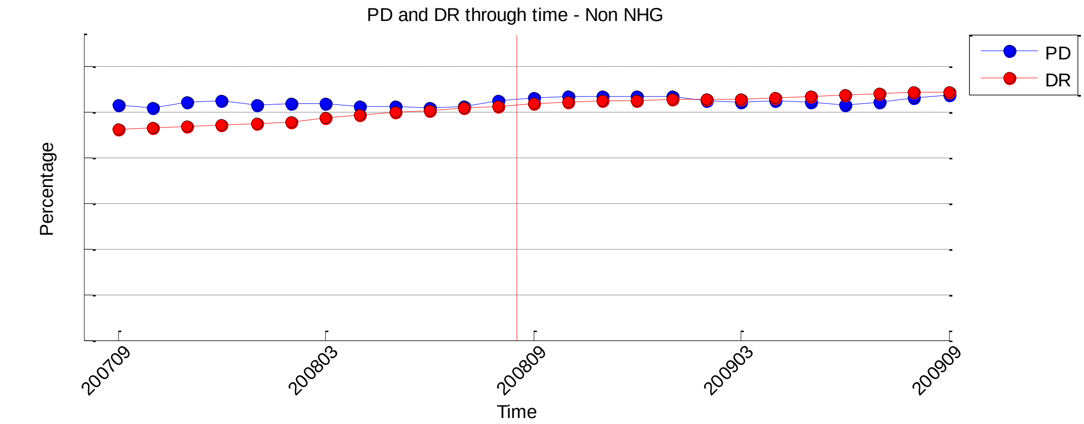


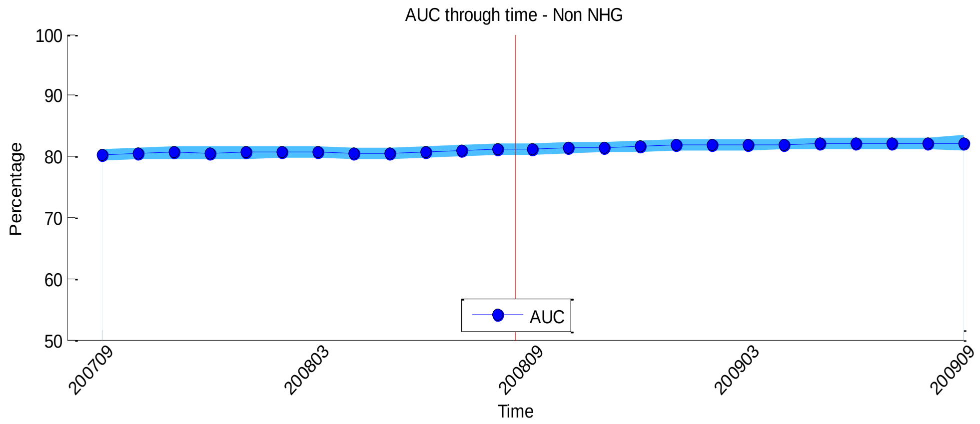
NHG/Non NHG

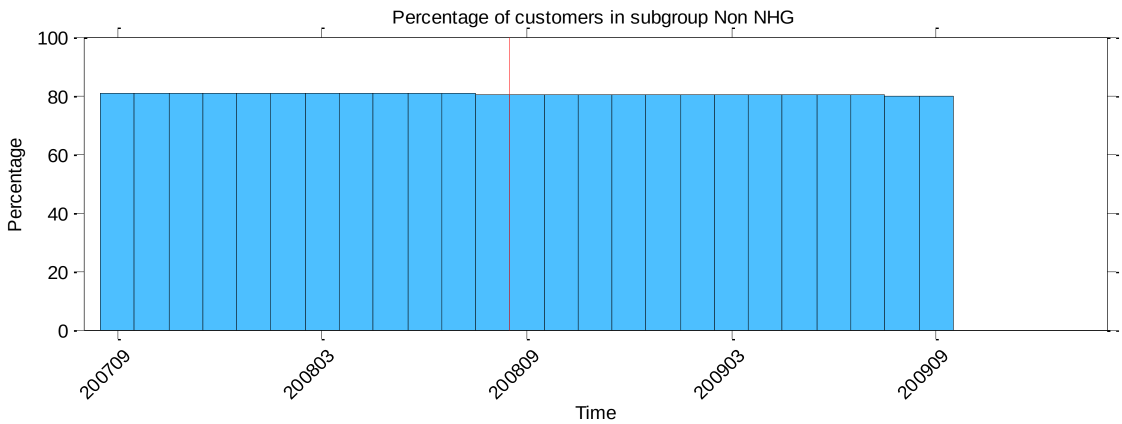

PD and DR through time - NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - Healthy
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

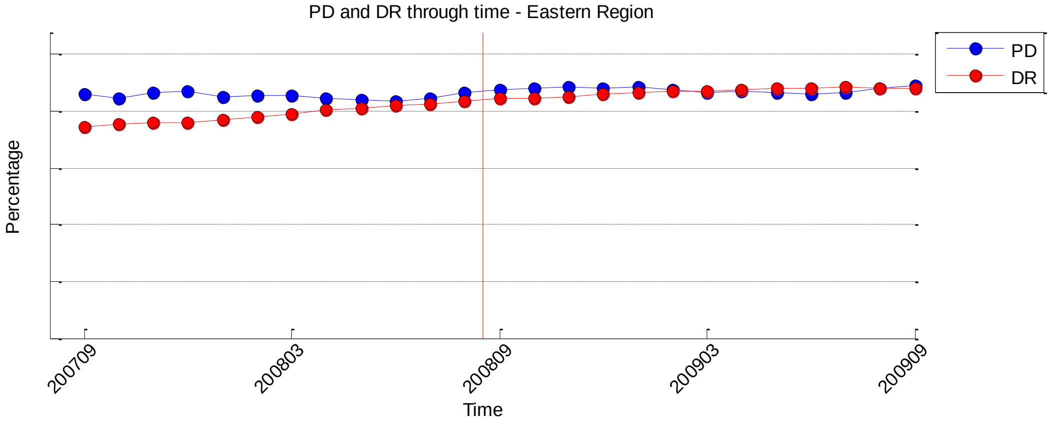

Percentage of customers in subgroup NHG
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

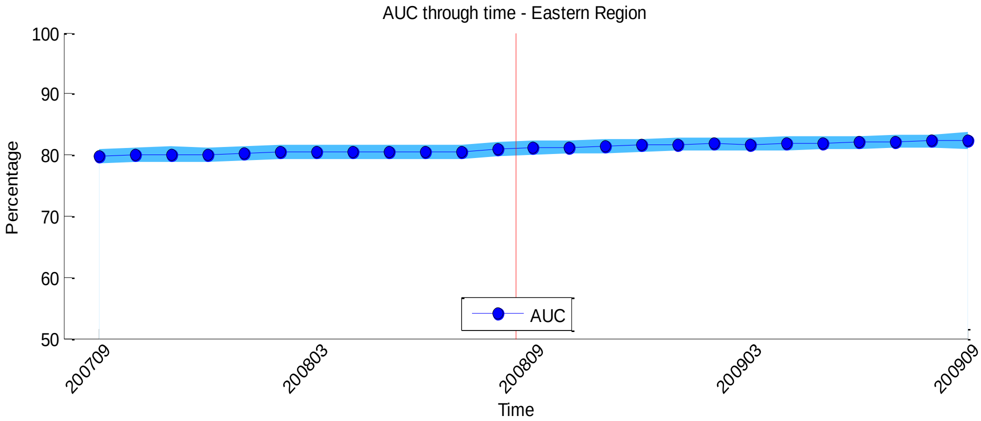

PD and DR through time - Non NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time

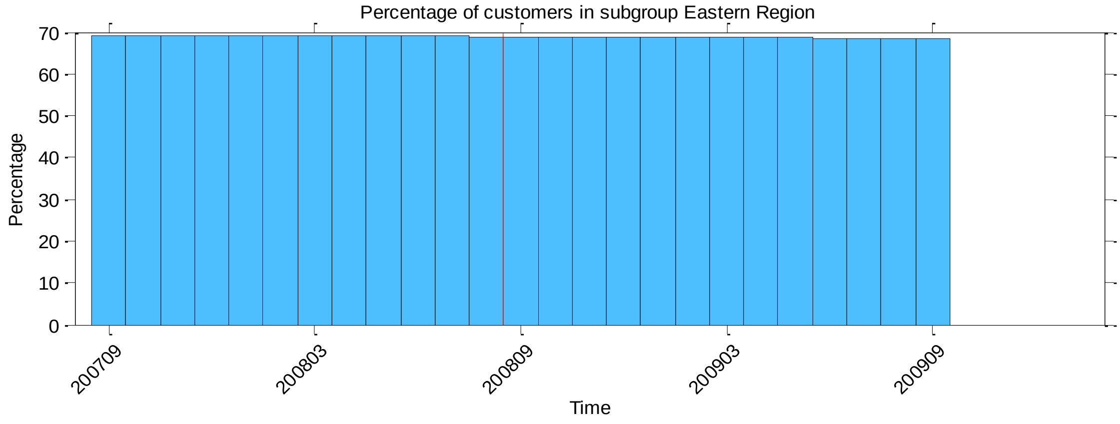

AUC through time - Non NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

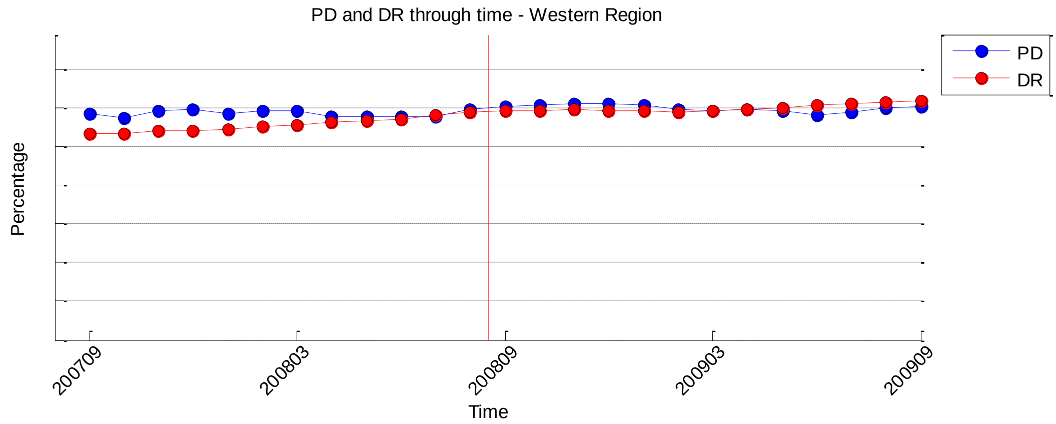

Percentage of customers in subgroup Non NHG
100
80
60
40
20
0
Percentage
200709
200803
200809
200903
200909
Time

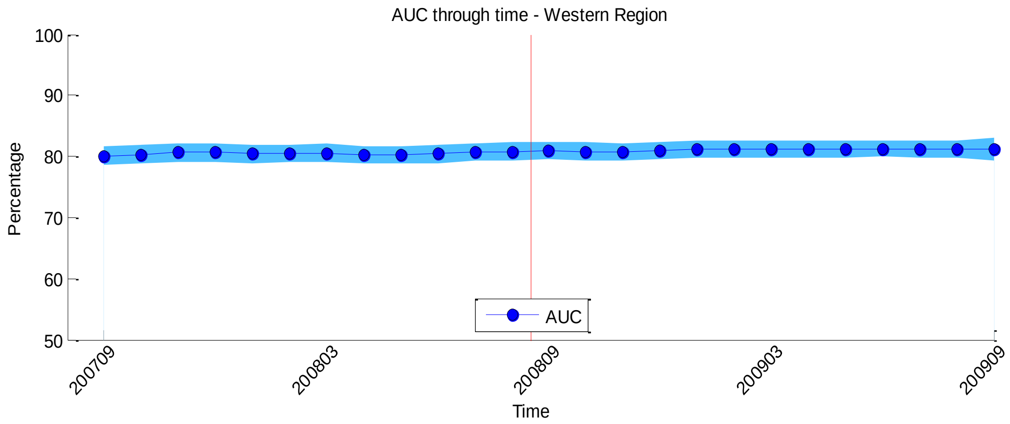
Eastern Region/Western Region

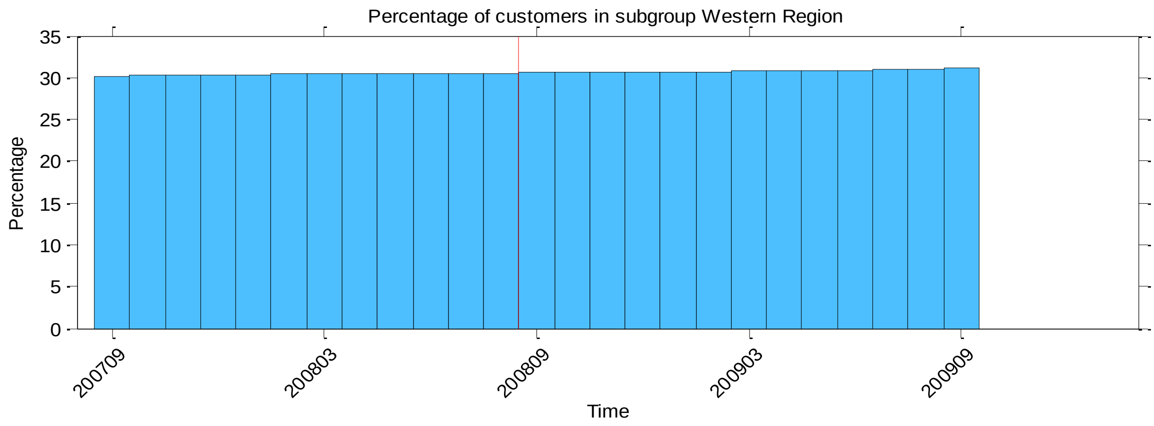






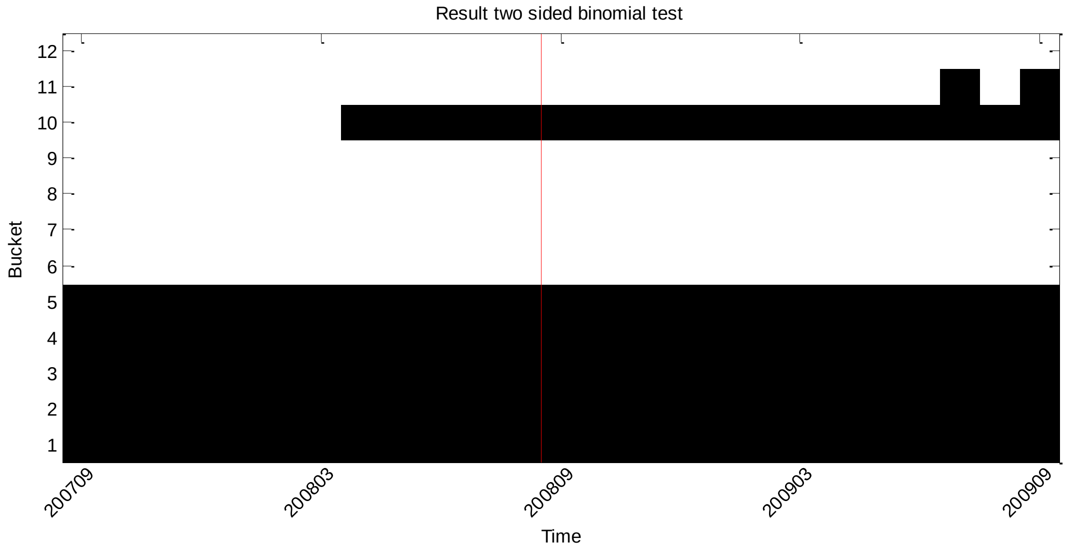

PD and DR through time - Western Region
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - Western Region
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

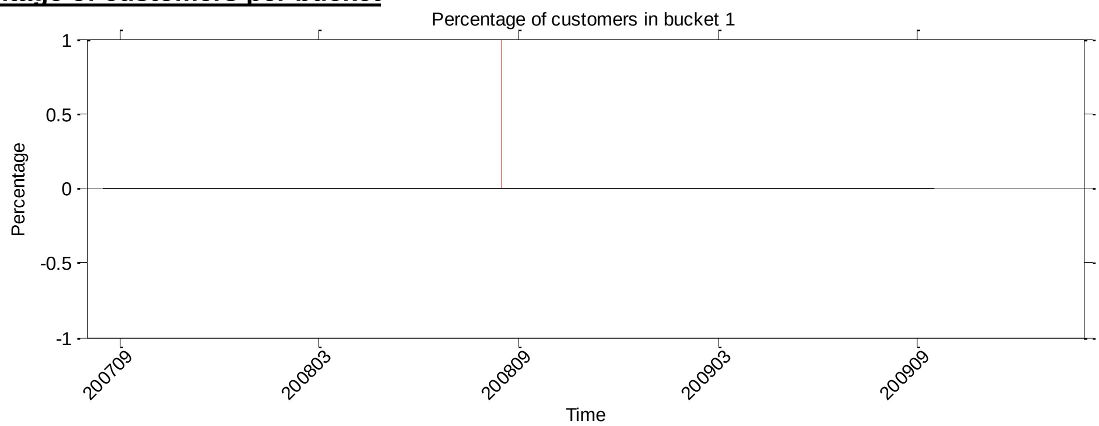

Percentage of customers in subgroup Western Region
35
30
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

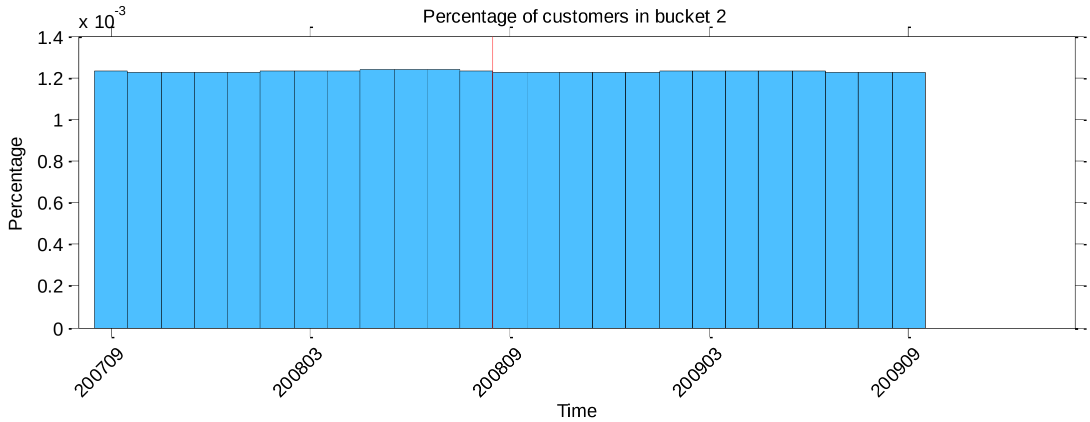
Binomial test Results on bucket level

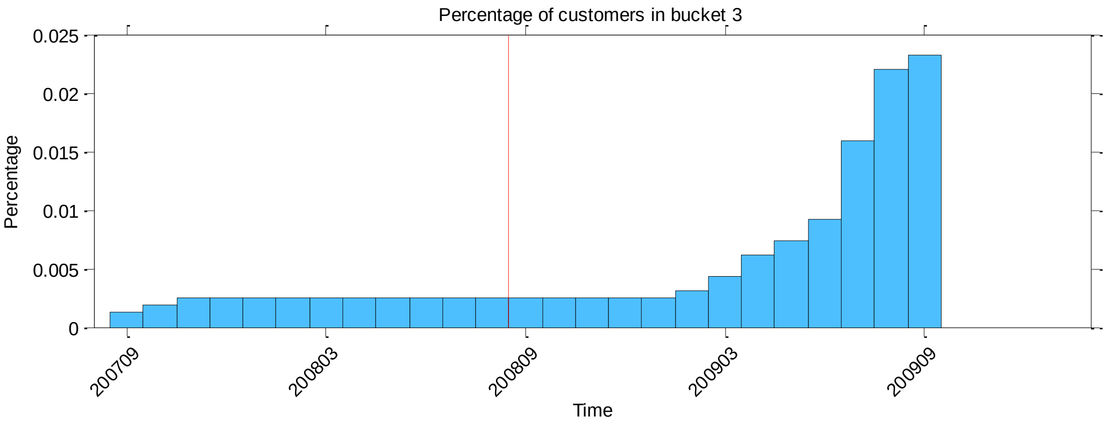
Important to notice is that the first 4 buckets contains less than 1 percent of the customers. This because a 5 year PD score is higher than a 1 year PD score. Because of the low number of observations in those buckets, the test results for those buckets cannot be seen as statistical relevant.

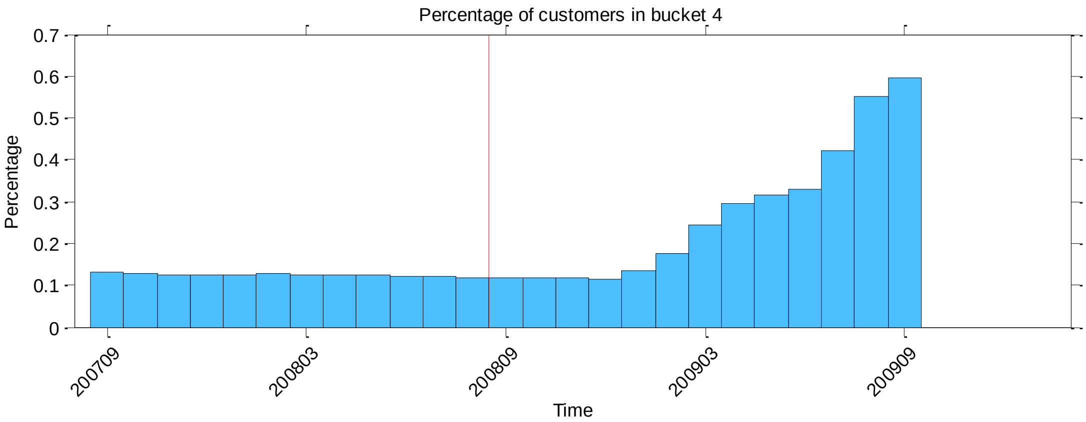


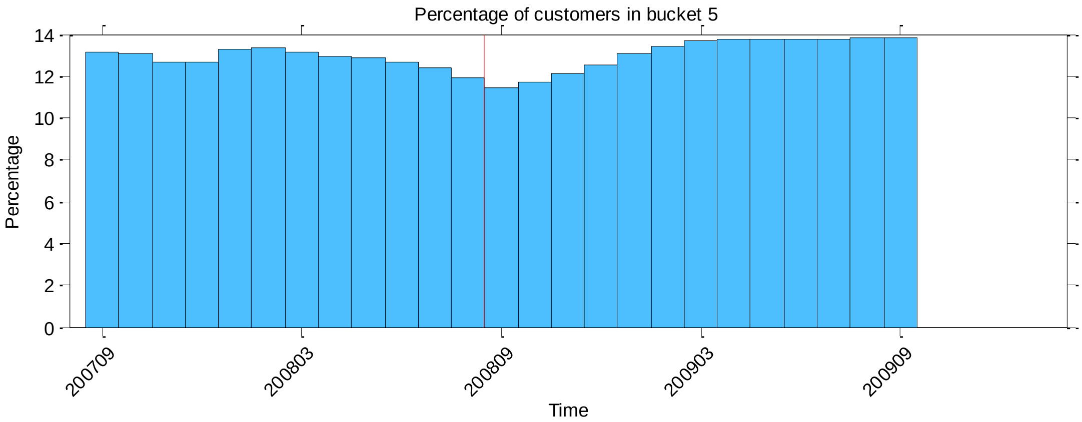
Percentage of customers per bucket

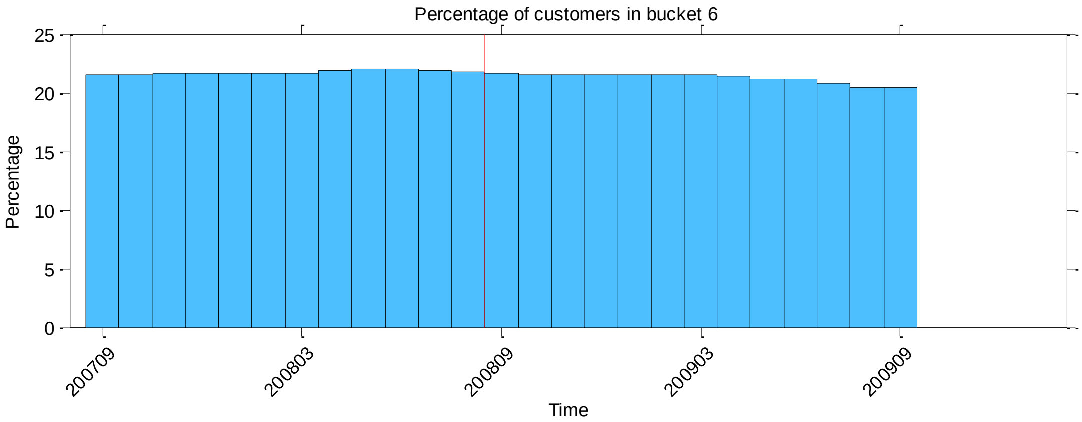


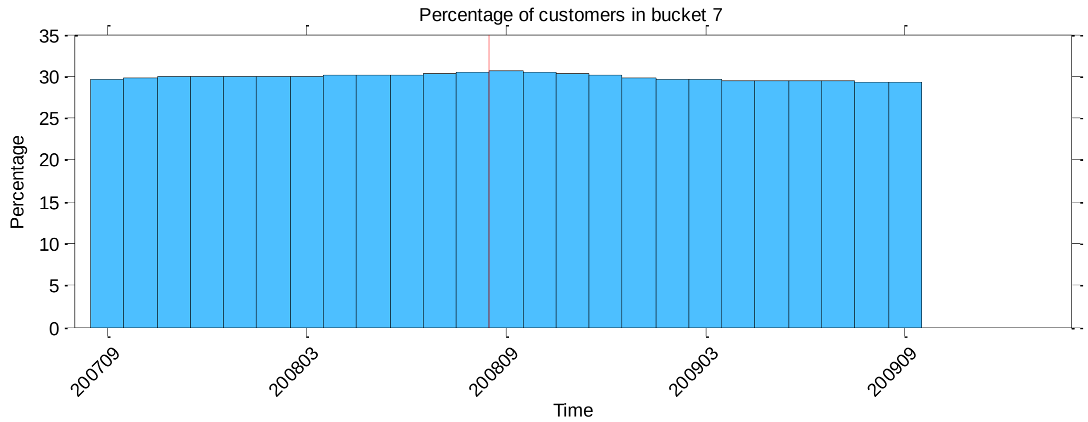


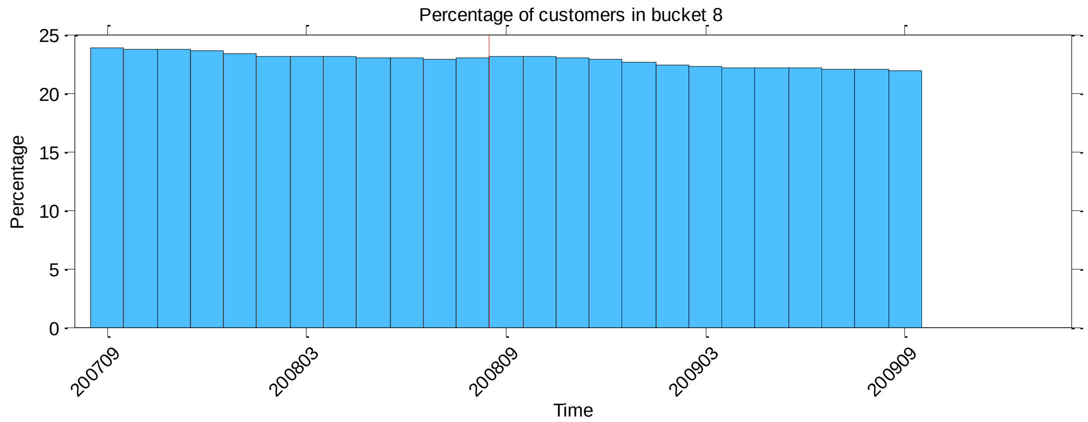

Percentage of customers in bucket 3
Percentage
0.025
0.02
0.015
0.01
0.005
0
200709
200803
200809
200903
200909
Time

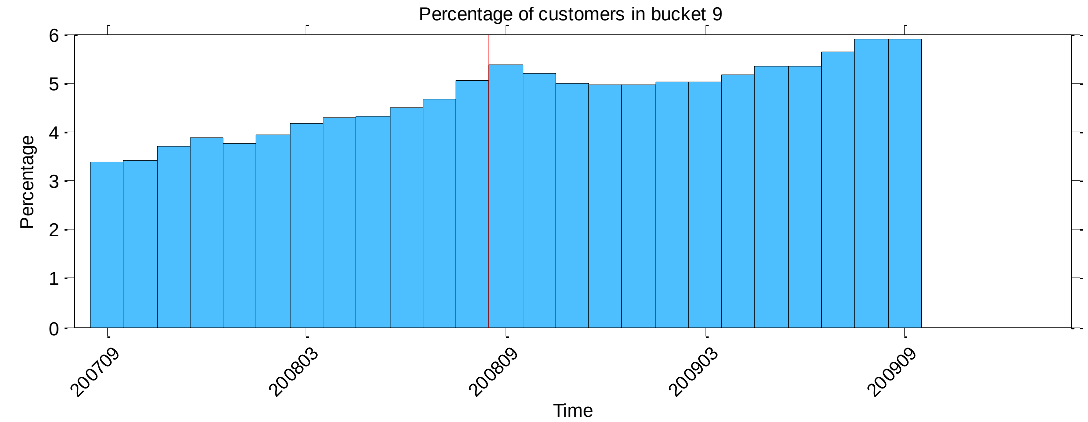

Percentage of customers in bucket 4
Percentage
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
200709
200803
200809
200903
200909
Time

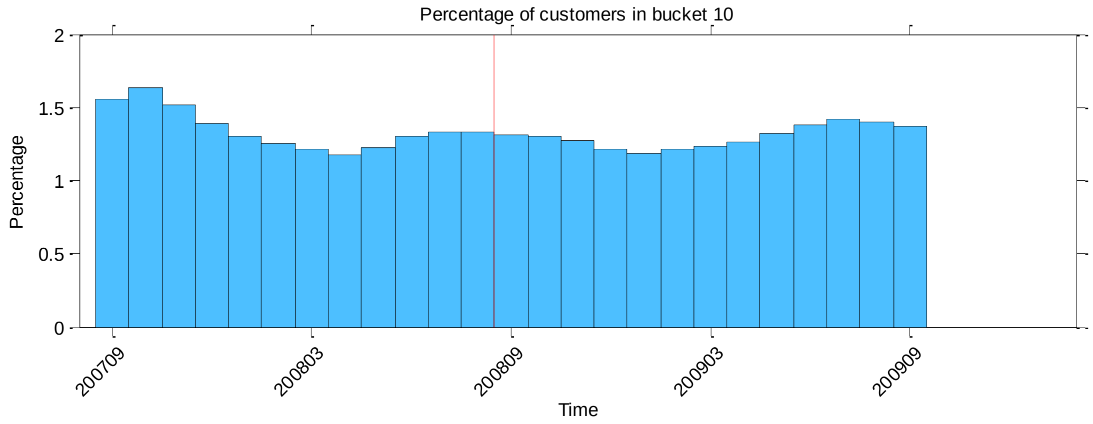

Percentage of customers in bucket 5
Percentage
14
12
10
8
6
4
2
0
200709
200803
200809
200903
200909
Time

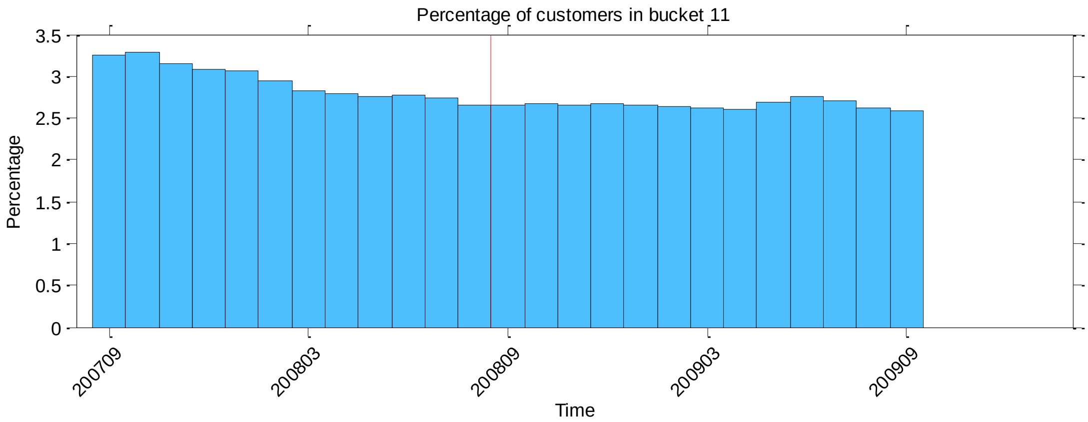

Percentage of customers in bucket 6
Percentage
25
20
15
10
5
0
200709
200803
200809
200903
200909
Time

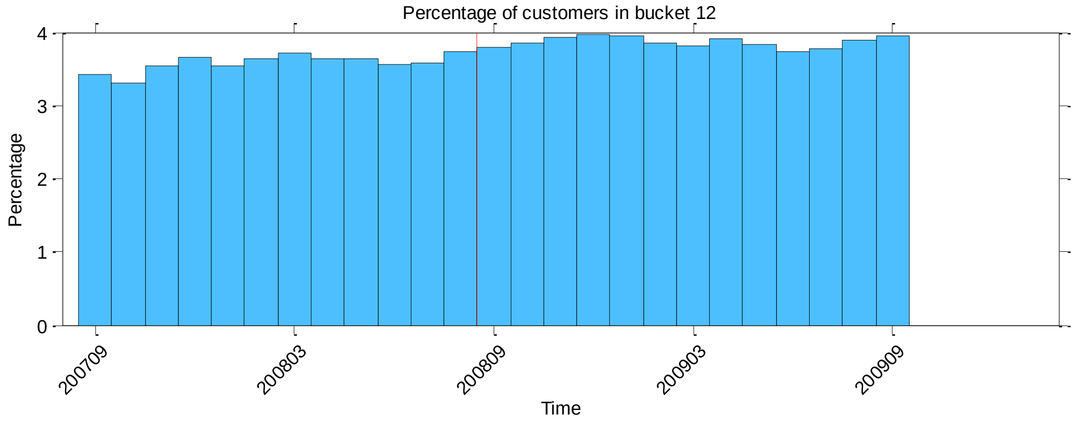

Percentage of customers in bucket 7
Percentage
Time
200709
200803
200809
200903
200909


Percentage of customers in bucket 8
Percentage
Time
200709
200803
200809
200903
200909


Percentage of customers in bucket 9
Percentage
Time
200709
200803
200809
200903
200909


Percentage of customers in bucket 10
Percentage
Time
200709
200803
200809
200903
200909

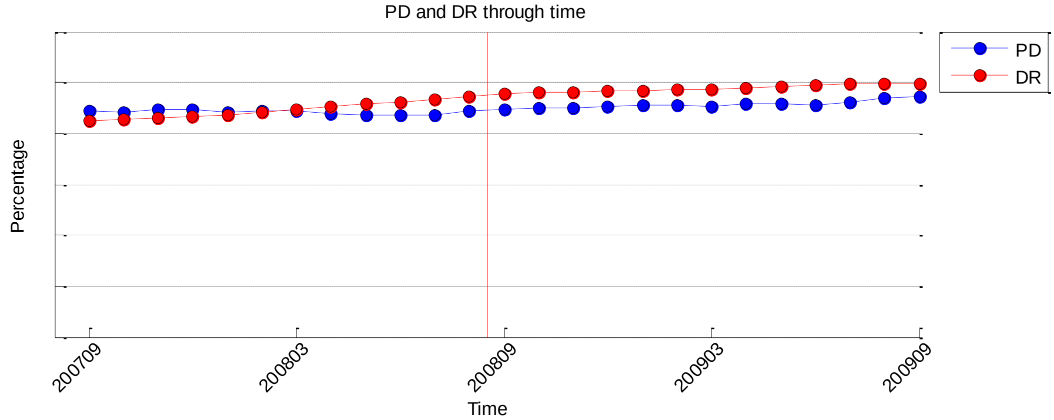

Percentage of customers in bucket 11
3.5
3
2.5
2
1.5
1
0.5
0
Percentage
200709
200803
200809
200903
200909
Time

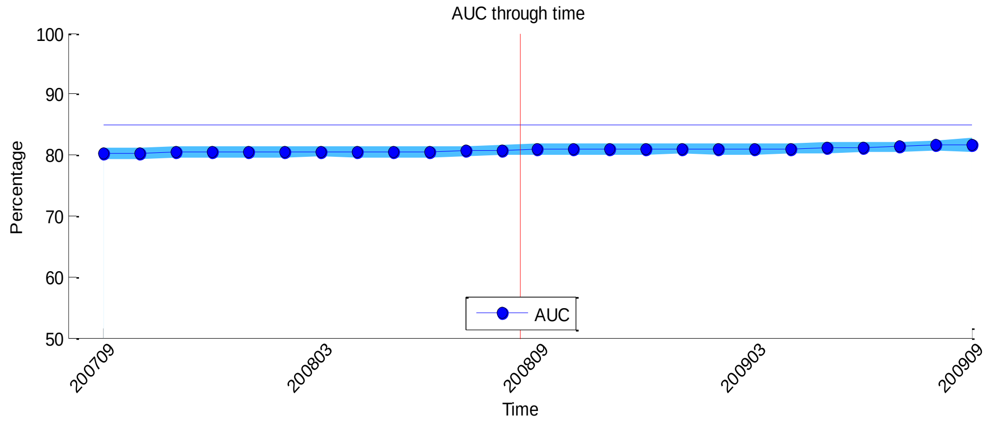

Percentage of customers in bucket 12
4
3
2
1
0
Percentage
200709
200803
200809
200903
200909
Time

APPENDIX 11 – Logit direct 5 year prototype results

This APPENDIX contains statistics and test results from the Logit LT prototype with mean Lambda extrapolation. The mean number of observations per moment of measurement is equal to 159381.56.

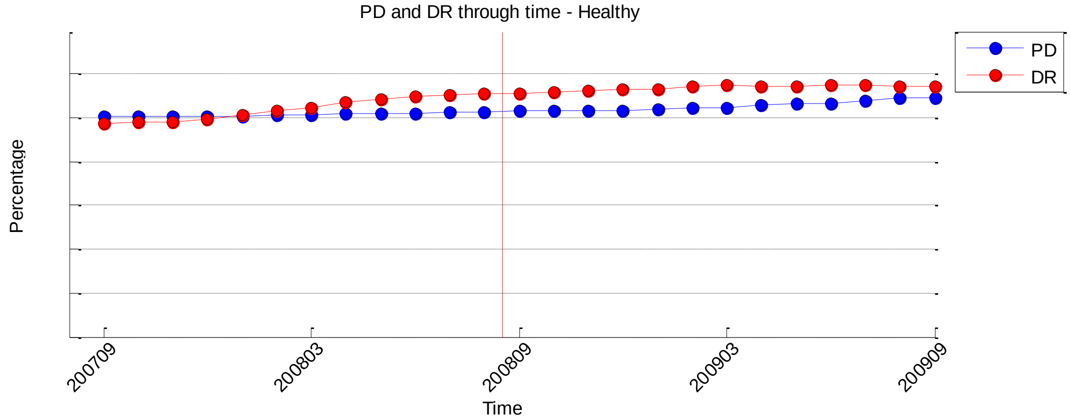
Portfolio

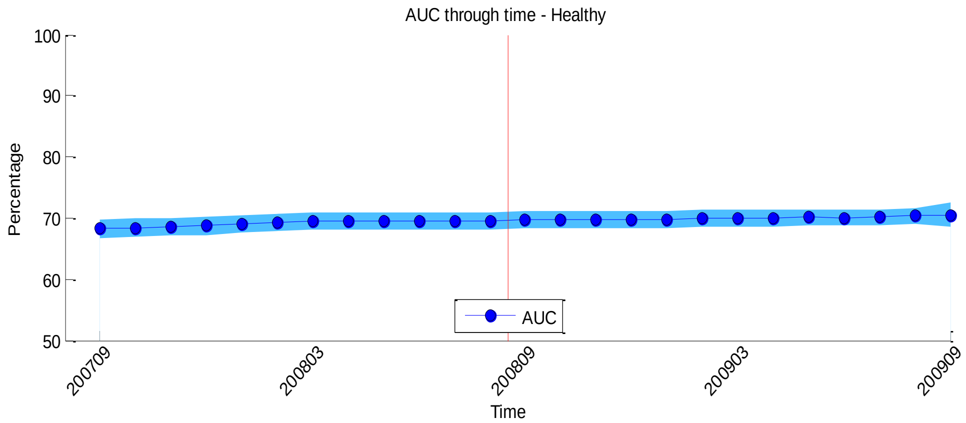


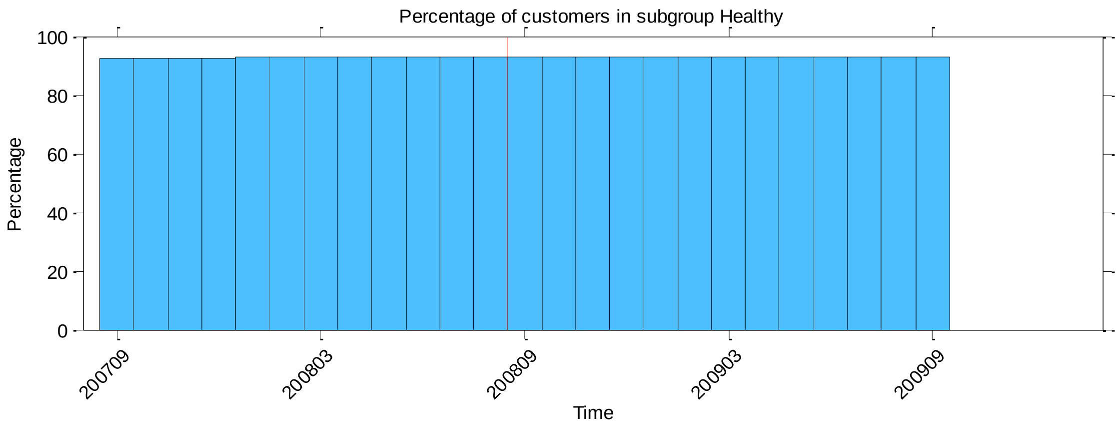


Healthy/Recovered/Default

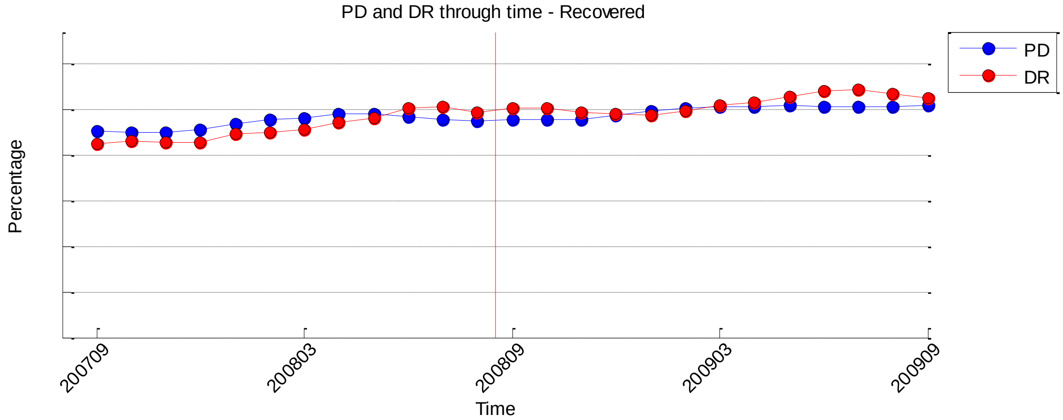
Healthy

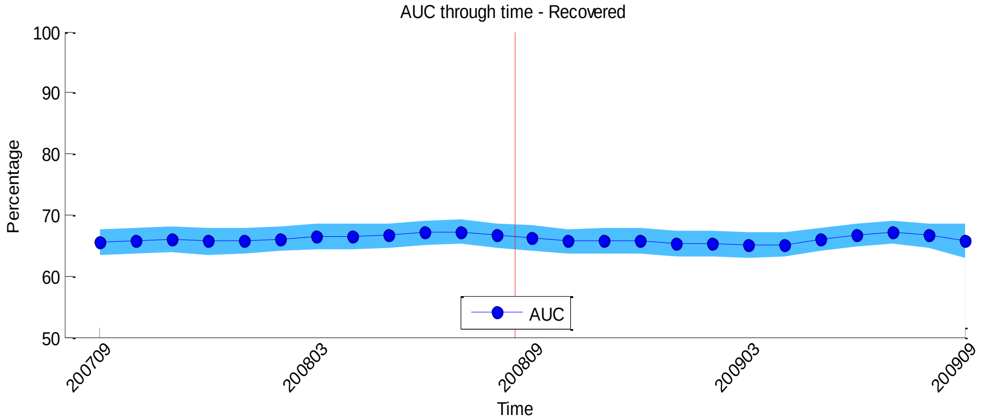


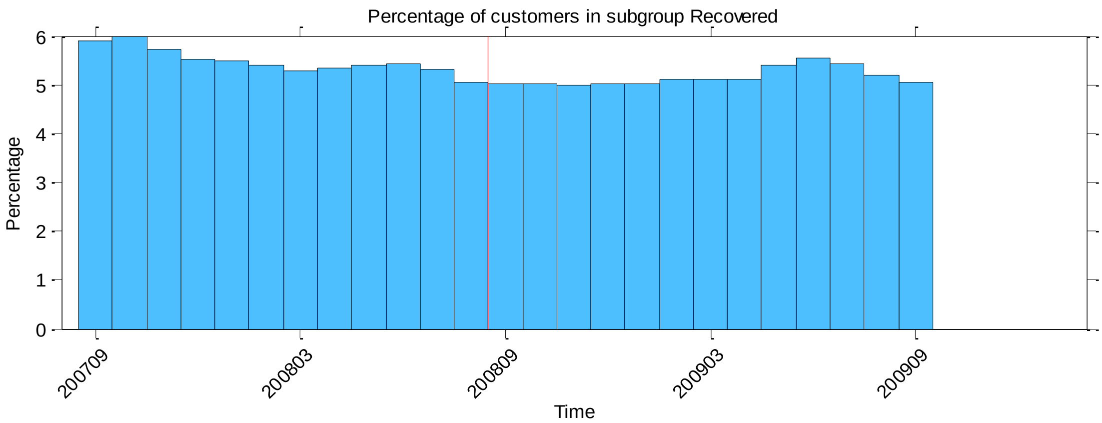




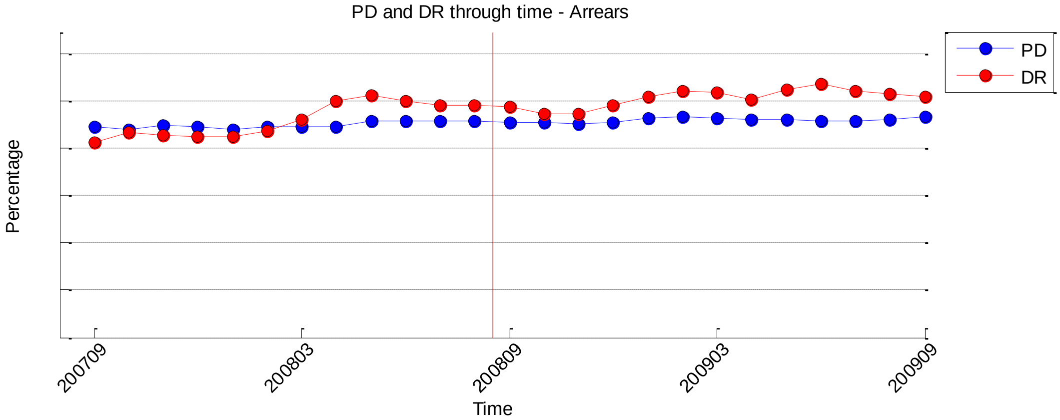
Recovered

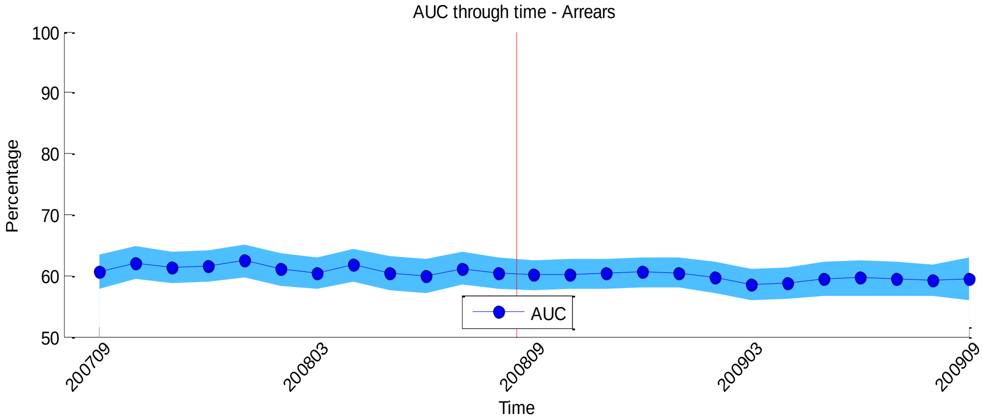

PD and DR through time - Recovered
PD
DR
Percentage
200709
200803
200809
200903
200909
Time

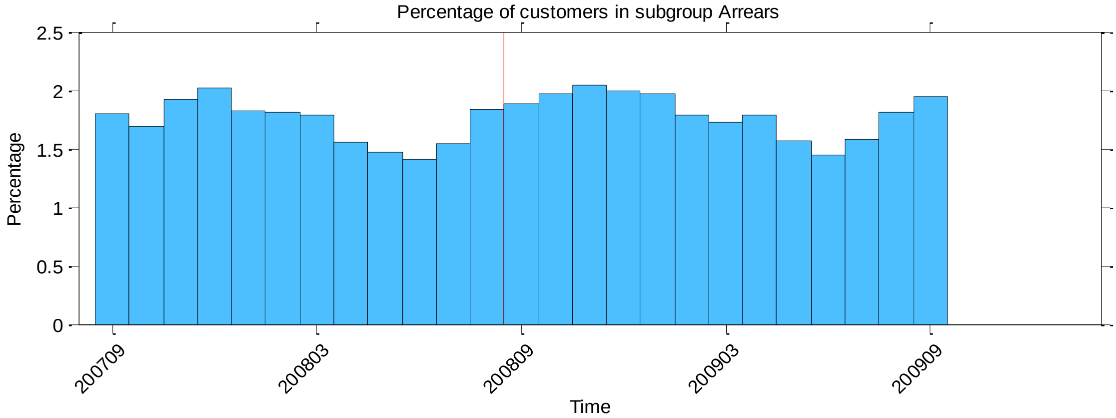

AUC through time - Recovered
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time


Percentage of customers in subgroup Recovered
6
5
4
3
2
1
0
Percentage
200709
200803
200809
200903
200909
Time

Arrears

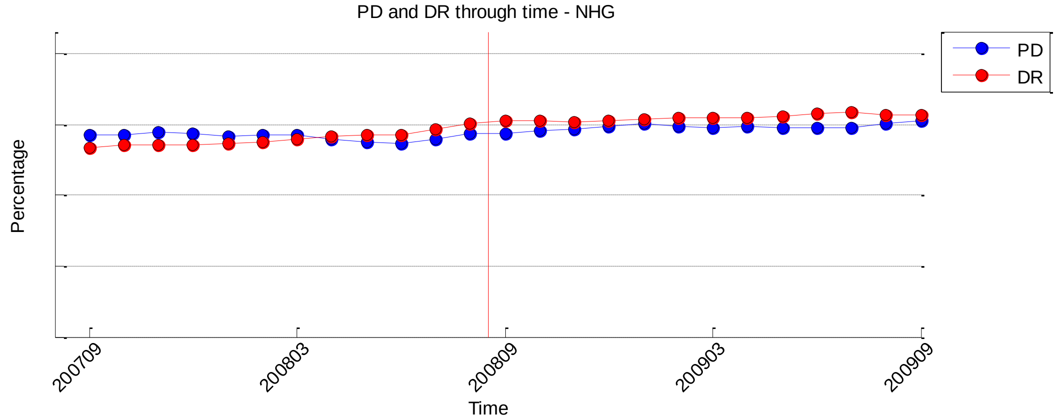

PD and DR through time - Arrears
PD
DR
Percentage
200709
200803
200809
200903
200909
Time

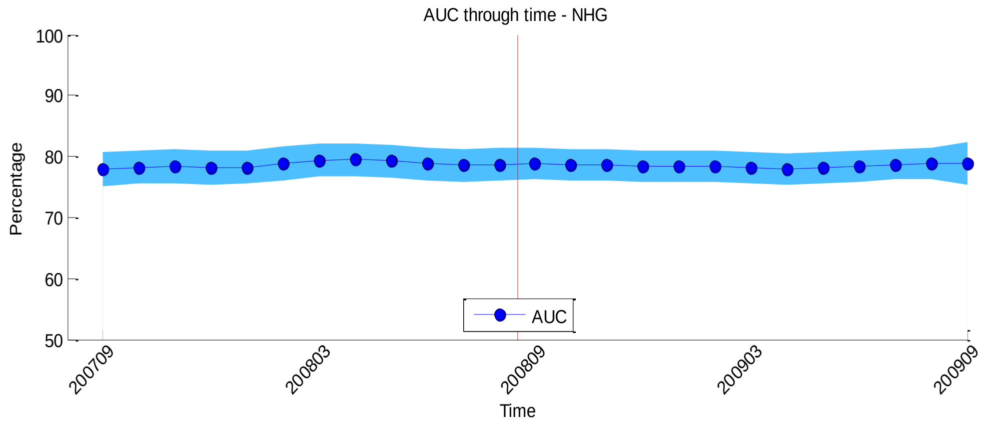

AUC through time - Arrears
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

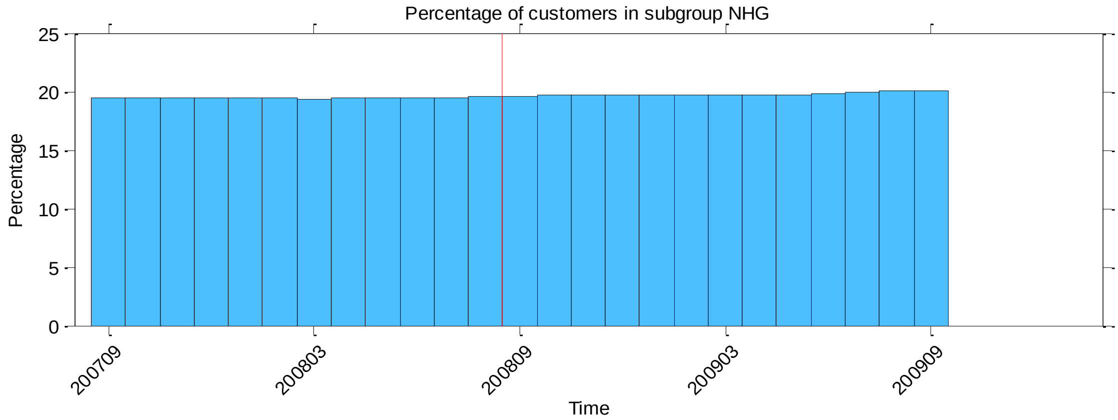

Percentage of customers in subgroup Arrears
2.5
2
1.5
1
0.5
0
Percentage
200709
200803
200809
200903
200909
Time

NHG/Non NHG

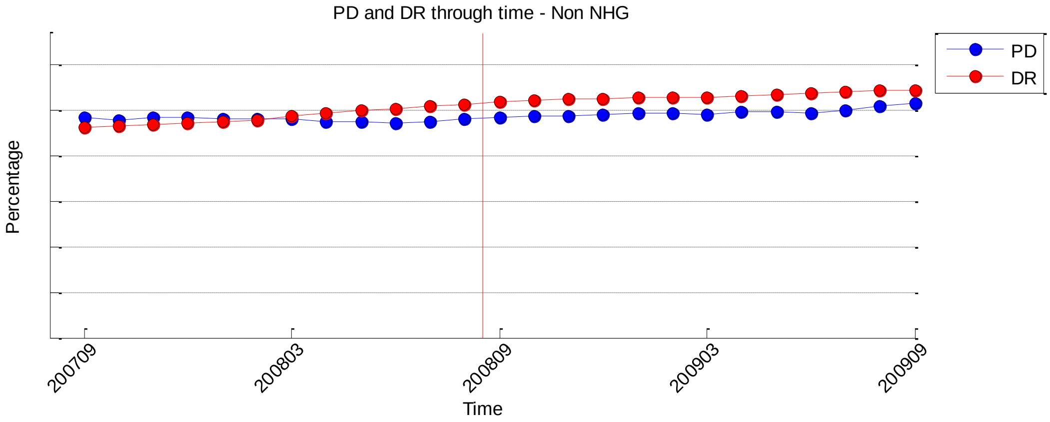
NHG

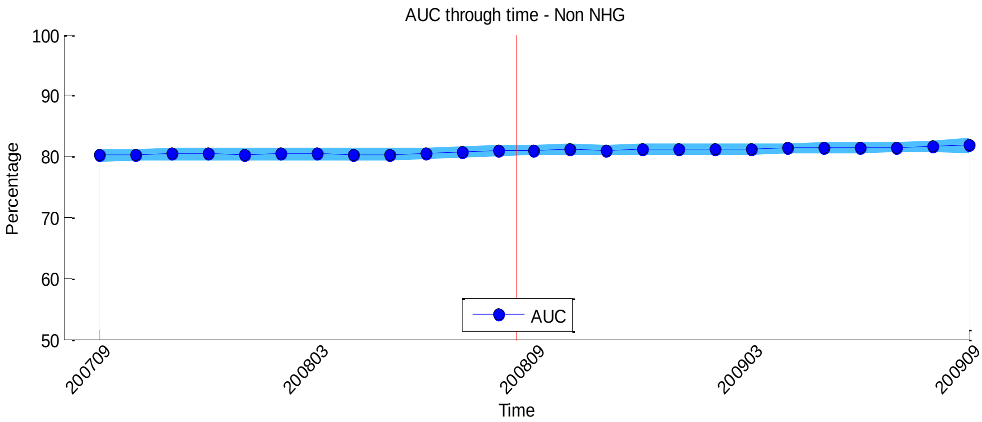


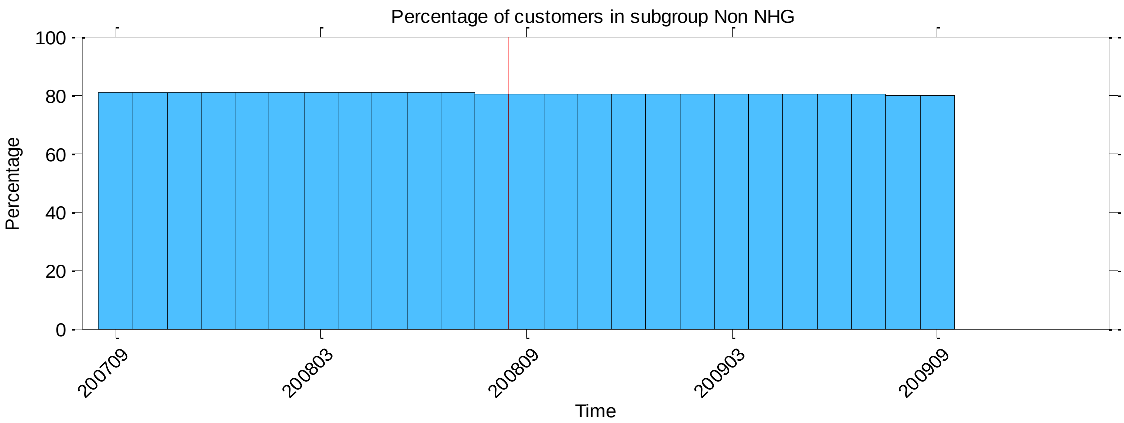




Non NHG

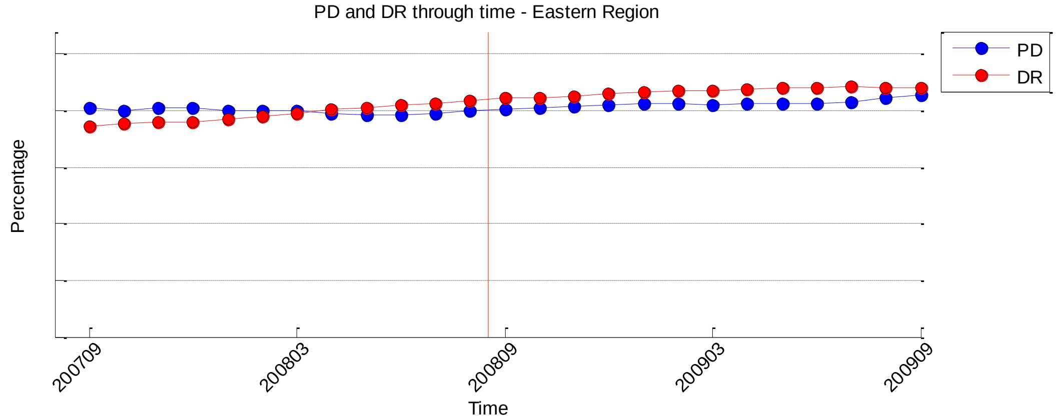

PD and DR through time - Non NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time

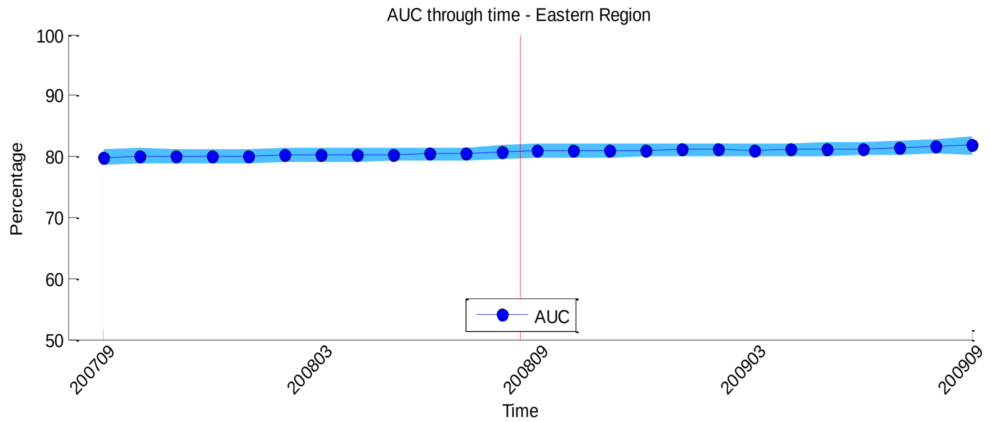

AUC through time - Non NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

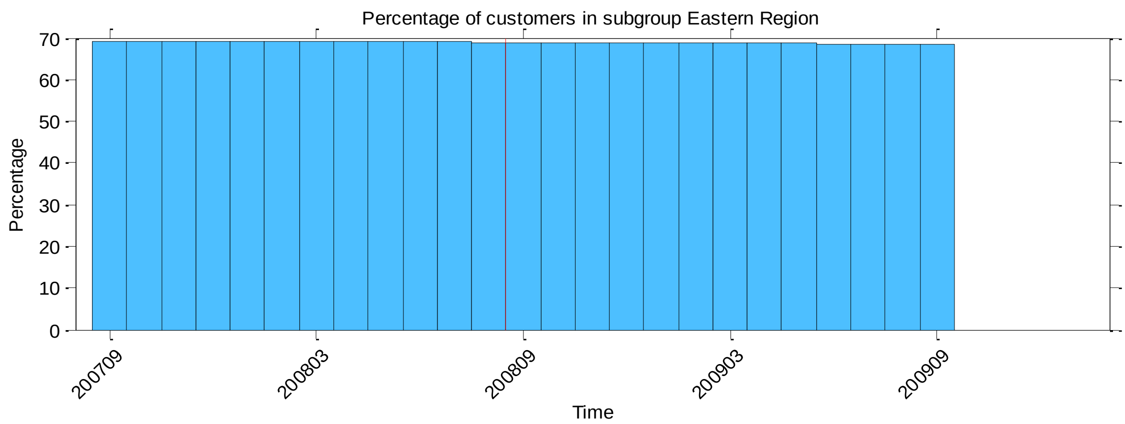

Percentage of customers in subgroup Non NHG
100
80
60
40
20
0
Percentage
200709
200803
200809
200903
200909
Time

Eastern Region/Western Region

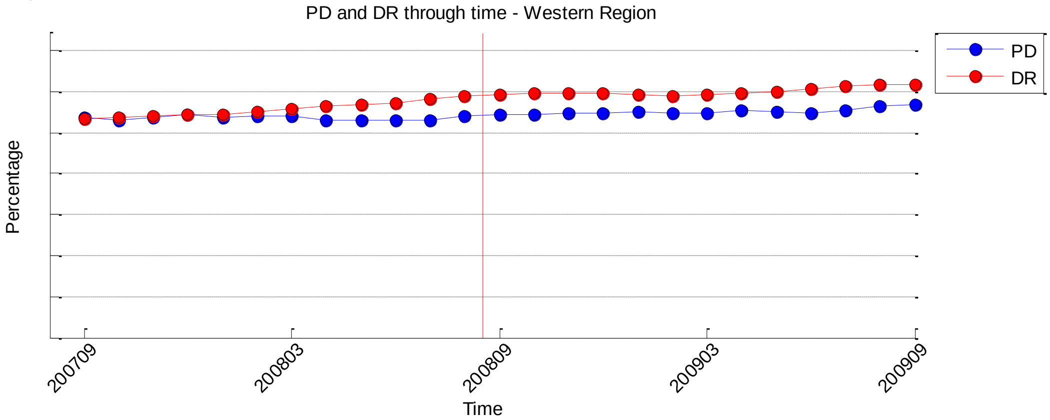
Eastern Region

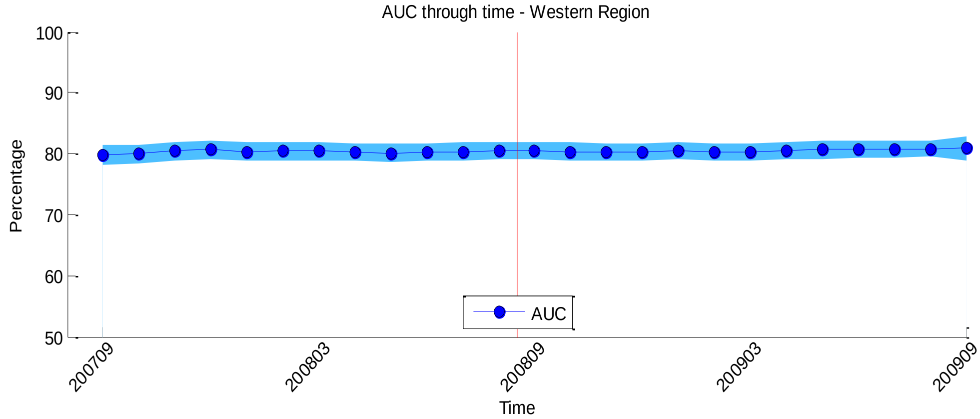


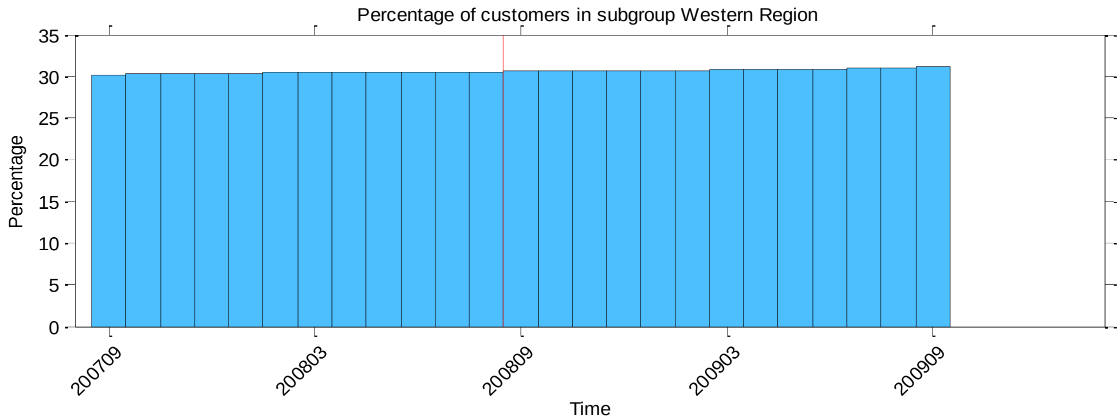




Western Region



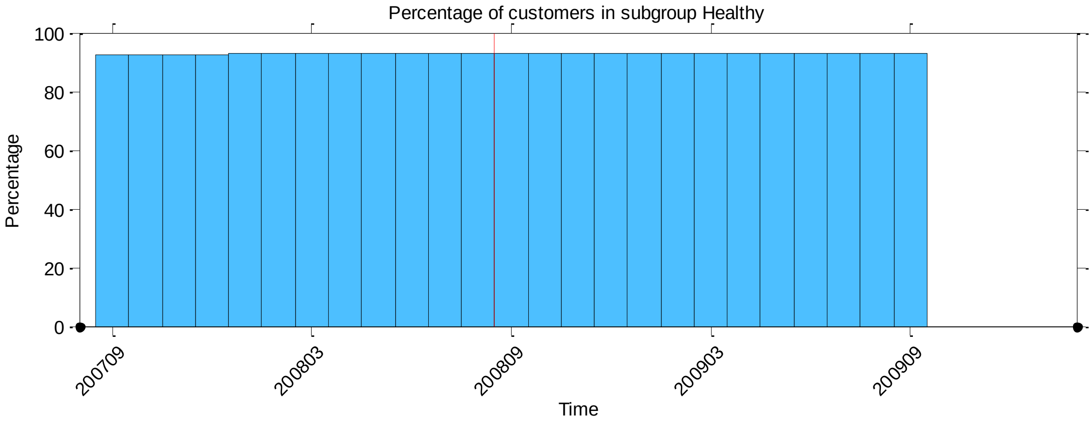


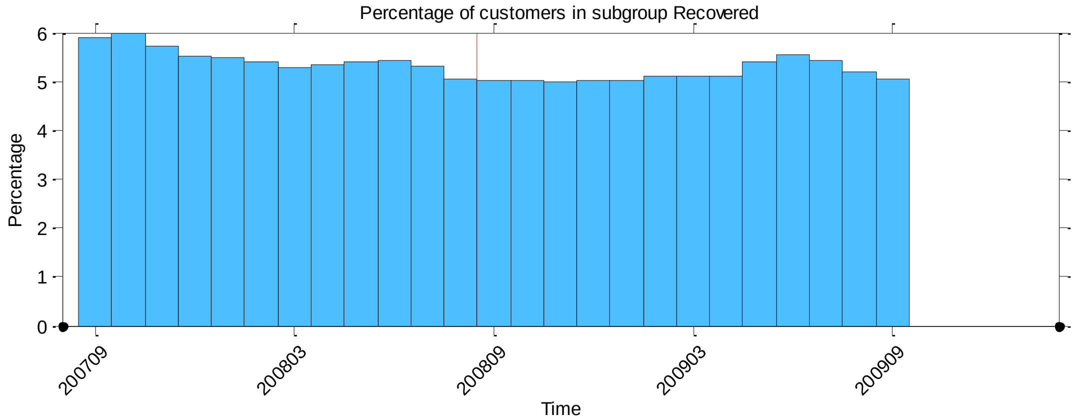


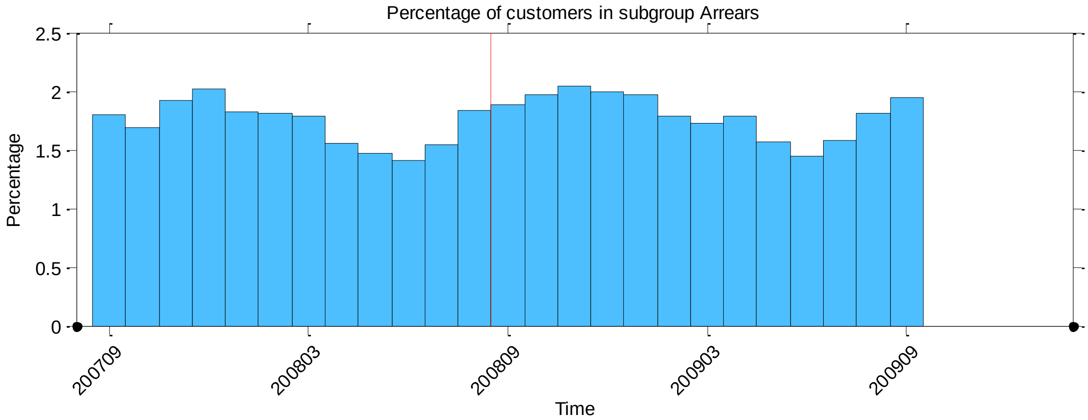
APPENDIX 12 – Logit LT prototype with Least Squares extrapolation results

This APPENDIX contains statistics and test results from the Logit LT prototype with mean Lambda extrapolation. The mean number of observations per moment of measurement is equal to 159381.56.

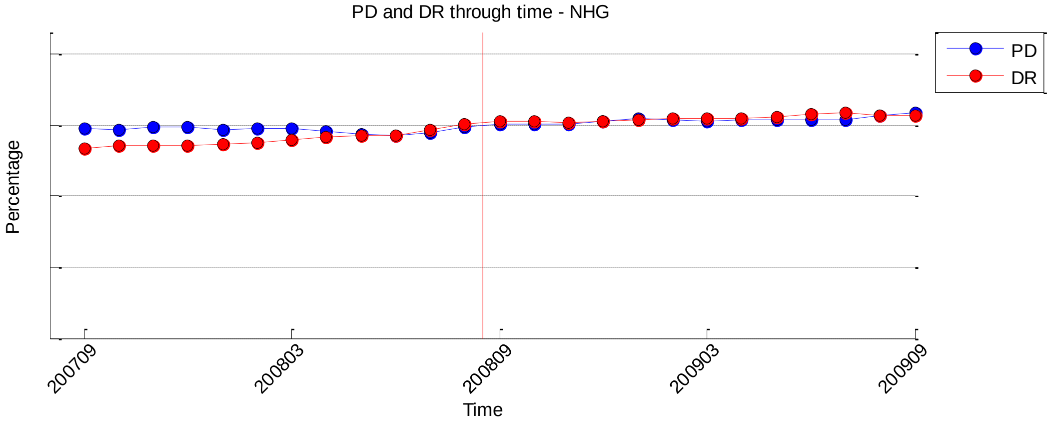
Healthy/Recovered/Default

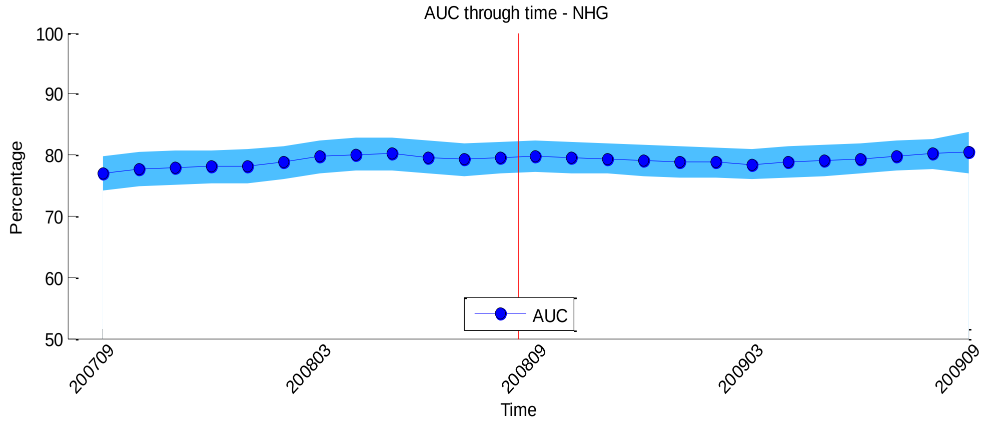


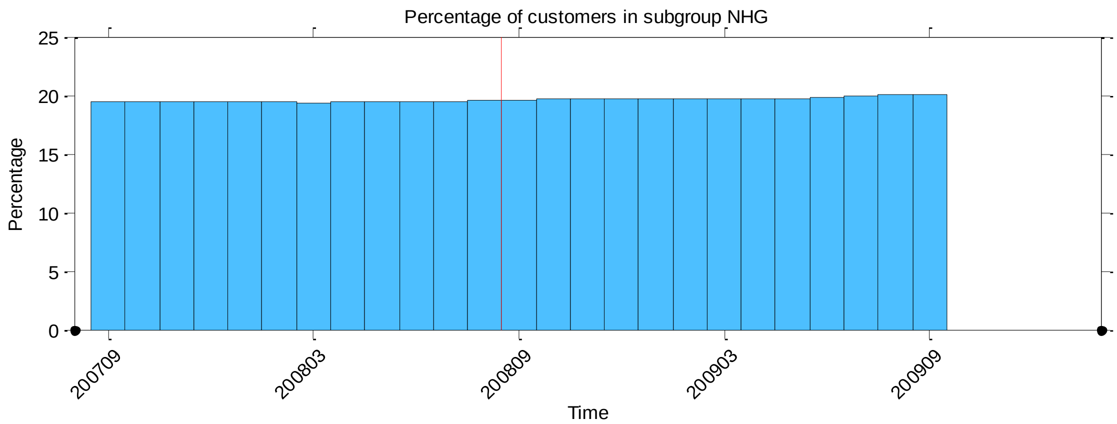


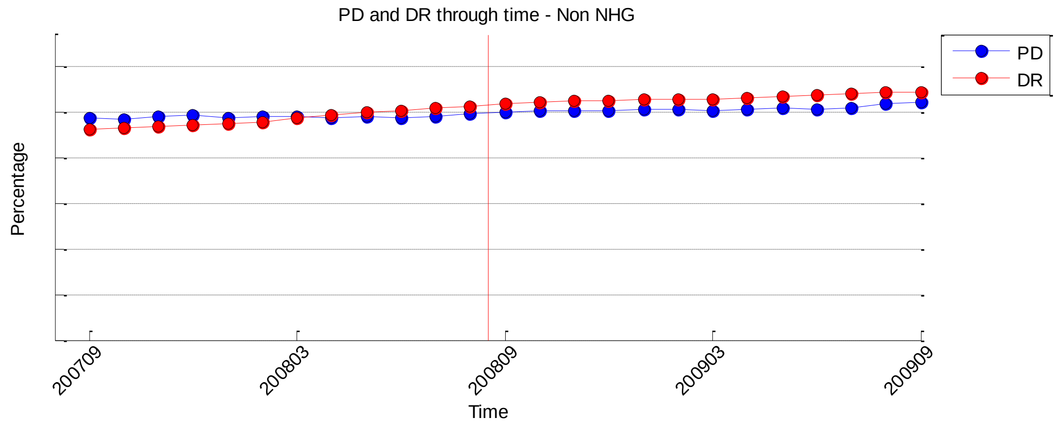


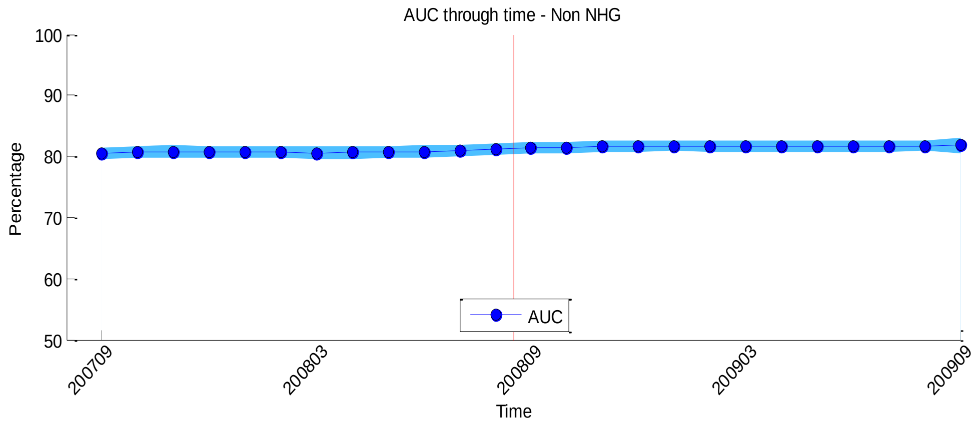
NHG/Non NHG

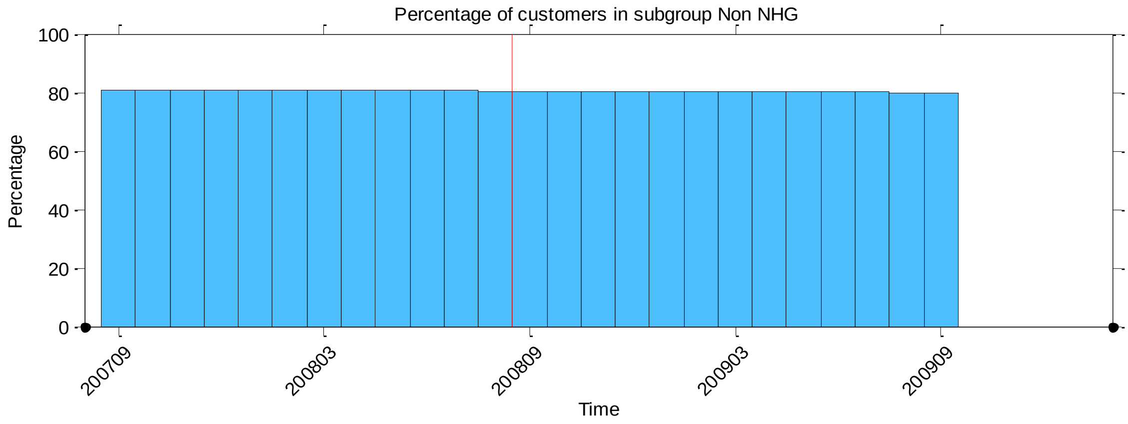

PD and DR through time - NHG
Percentage
PD
DR
200709
200803
200809
200903
200909
Time


AUC through time - NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

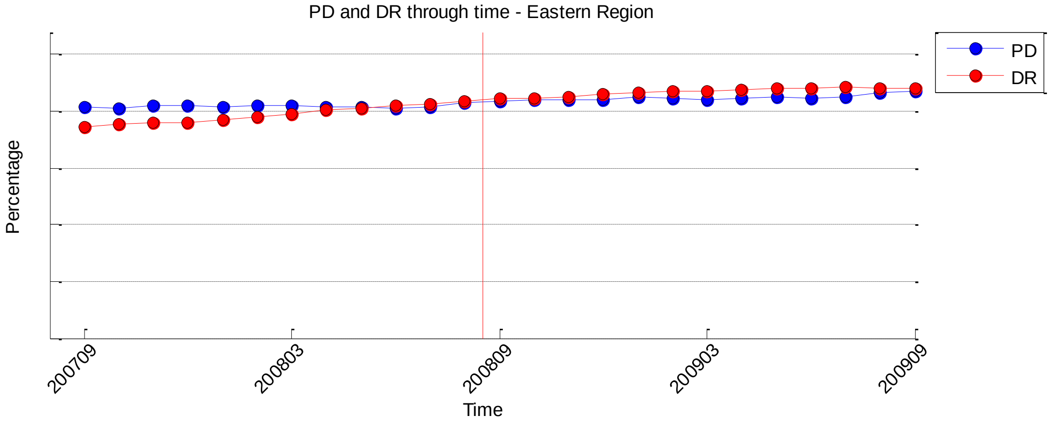

Percentage of customers in subgroup NHG
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

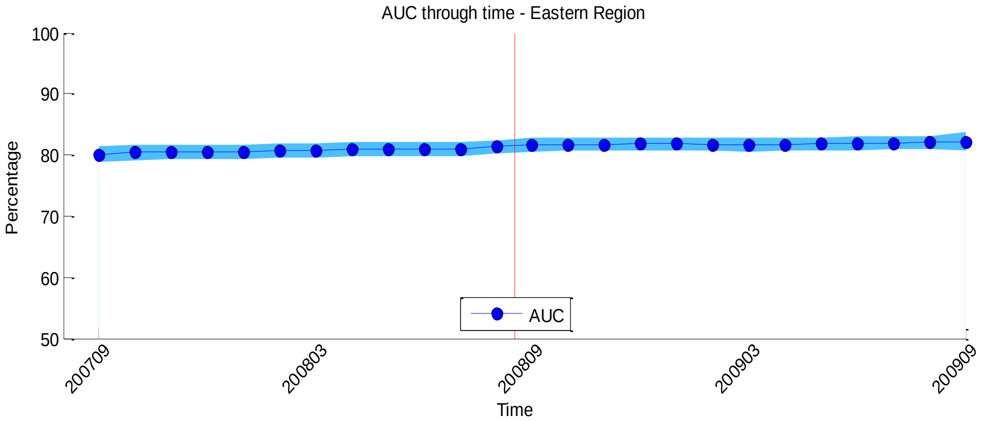

PD and DR through time - Non NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time

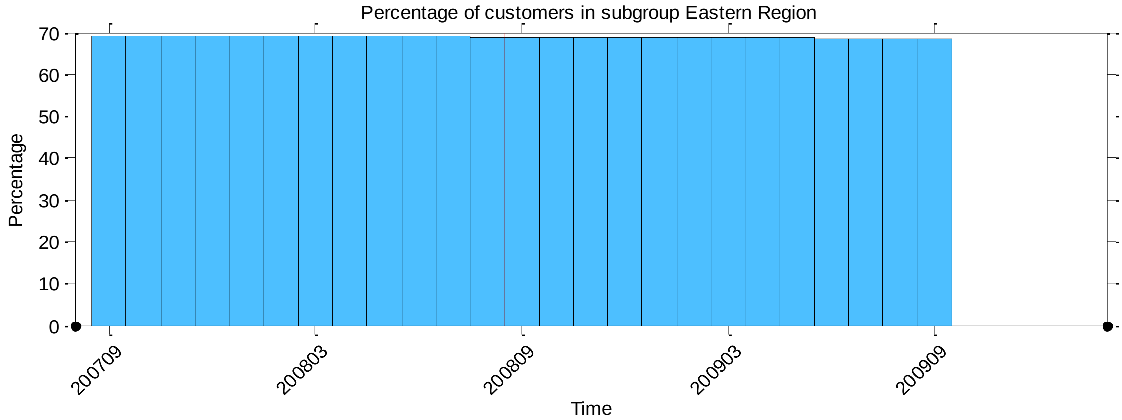

AUC through time - Non NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

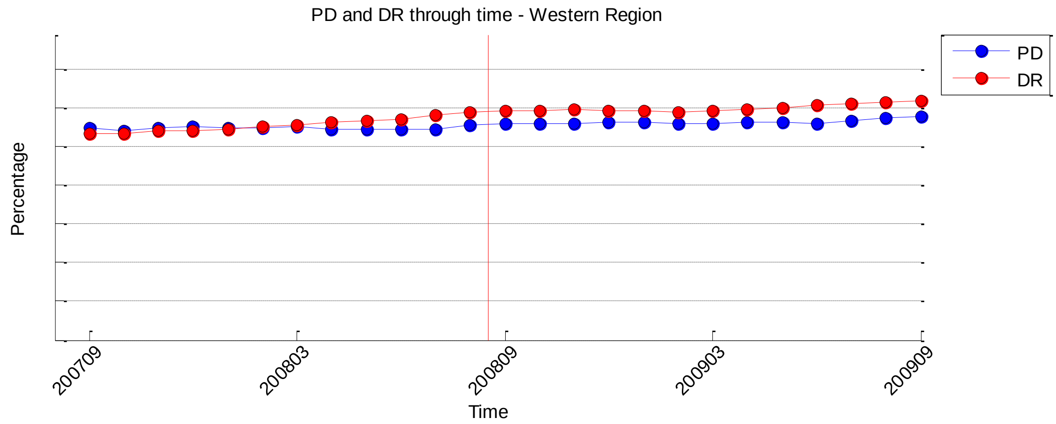

Percentage of customers in subgroup Non NHG
100
80
60
40
20
0
Percentage
200709
200803
200809
200903
200909
Time

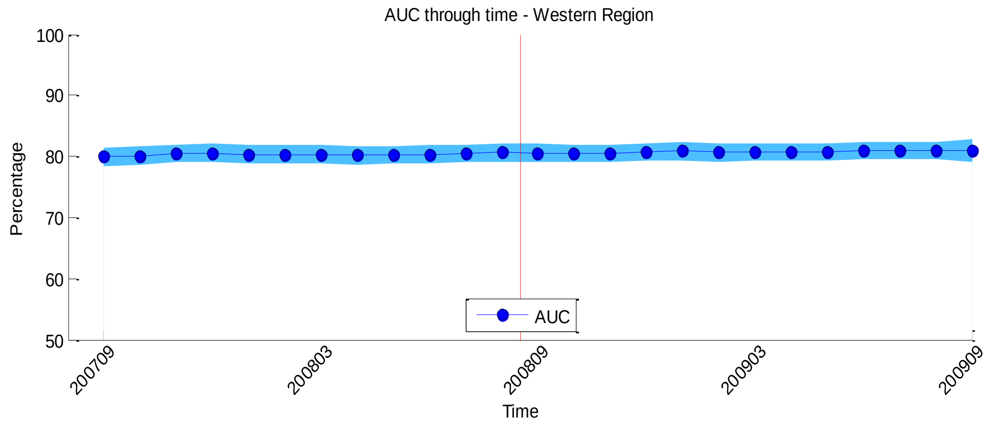
Eastern Region/Western Region

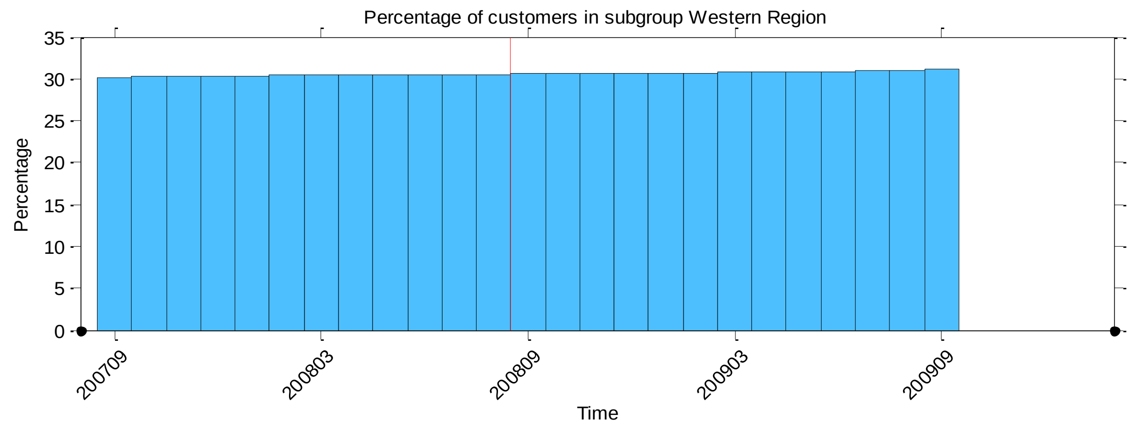






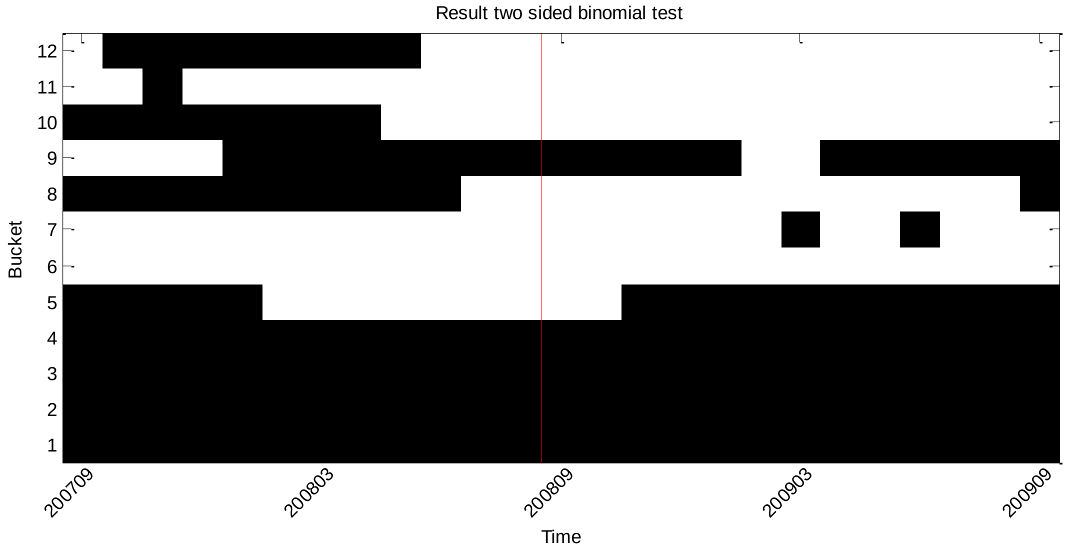

PD and DR through time - Western Region
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - Western Region
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

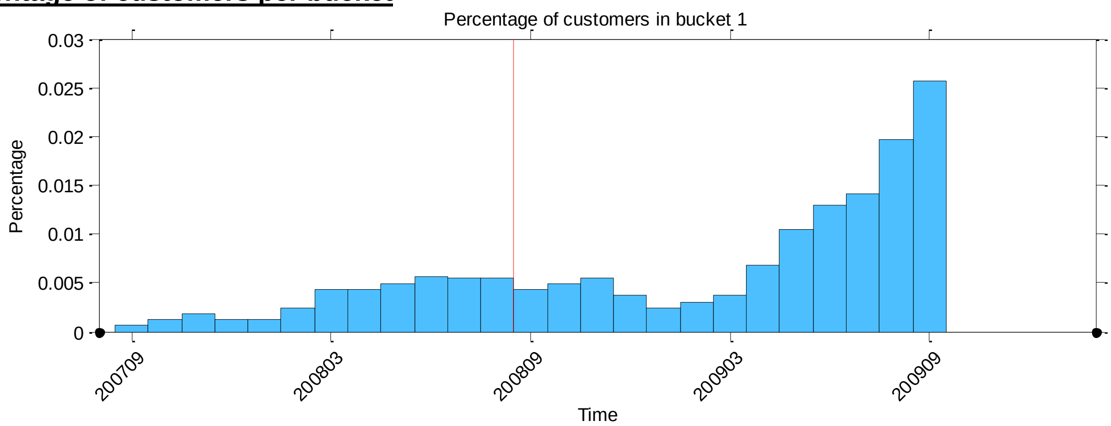

Percentage of customers in subgroup Western Region
35
30
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

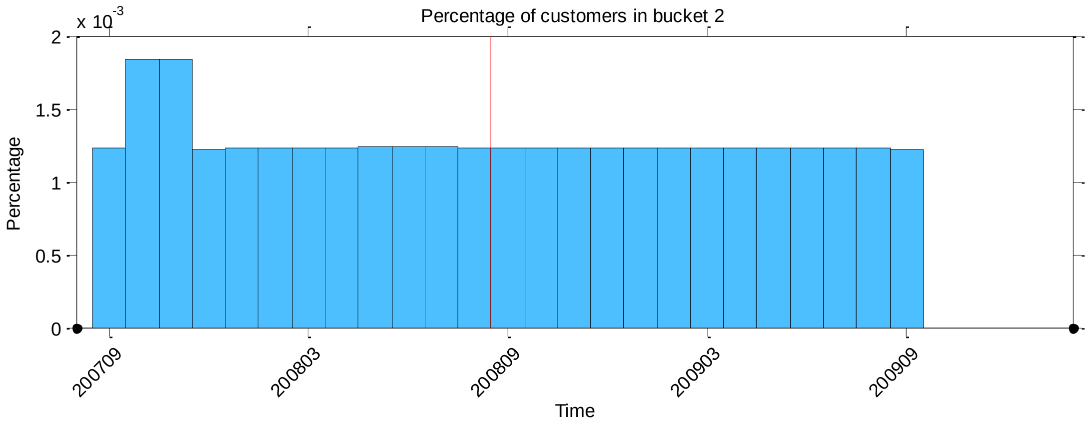
Binomial test Results on bucket level

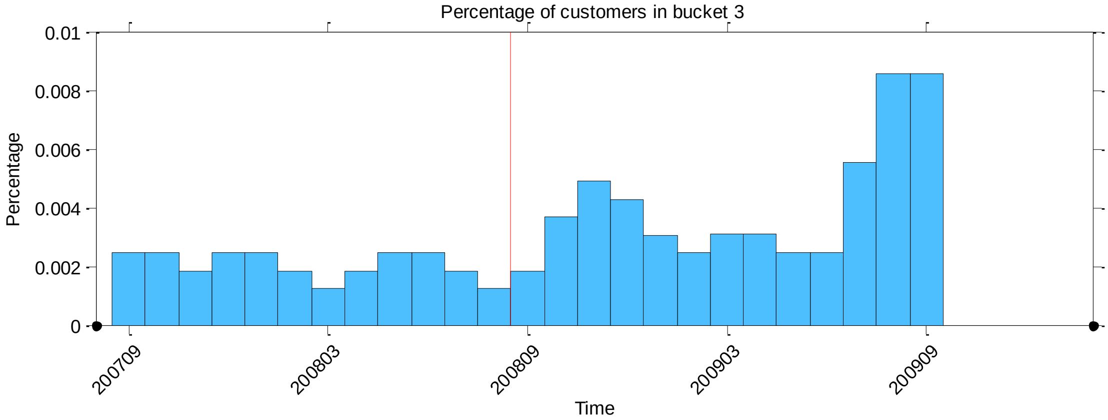
Important to notice is that the first 4 buckets contains less than 1 percent of the customers. This because a 5 year PD score is higher than a 1 year PD score. Because of the low number of observations in those buckets, the test results for those buckets cannot be seen as statistical relevant.

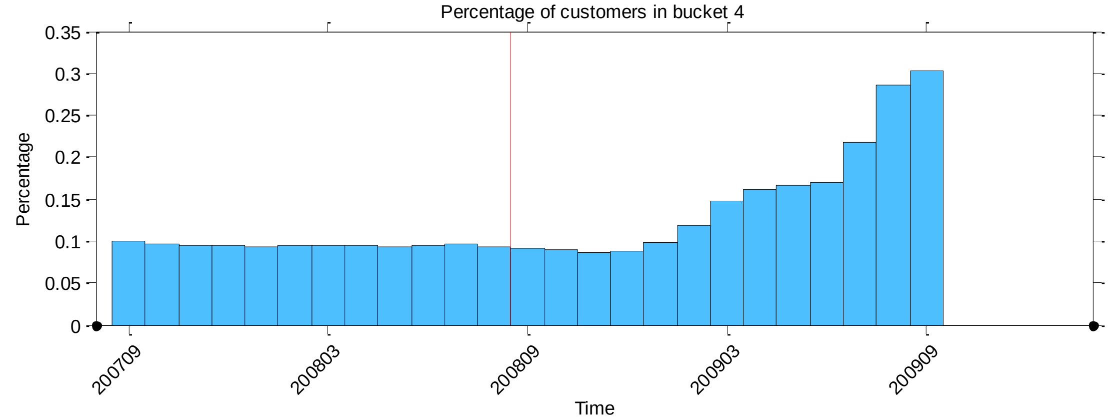


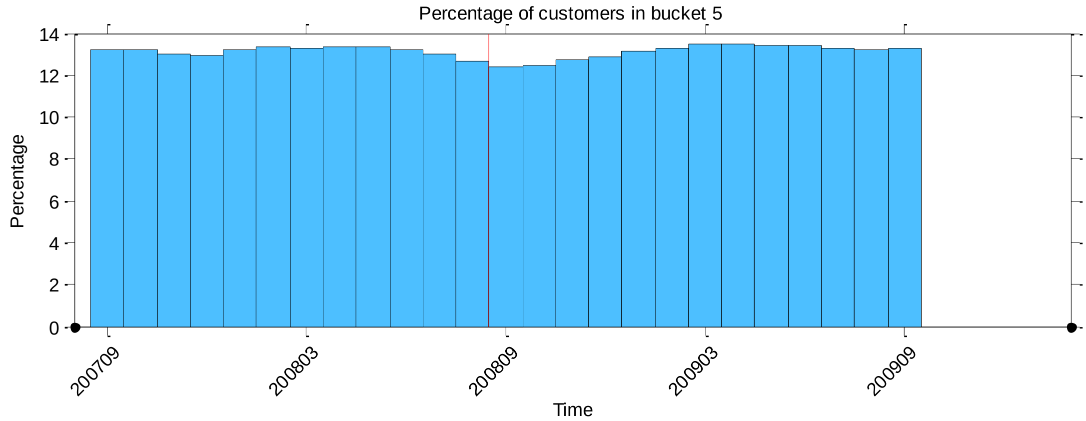
Percentage of customers per bucket

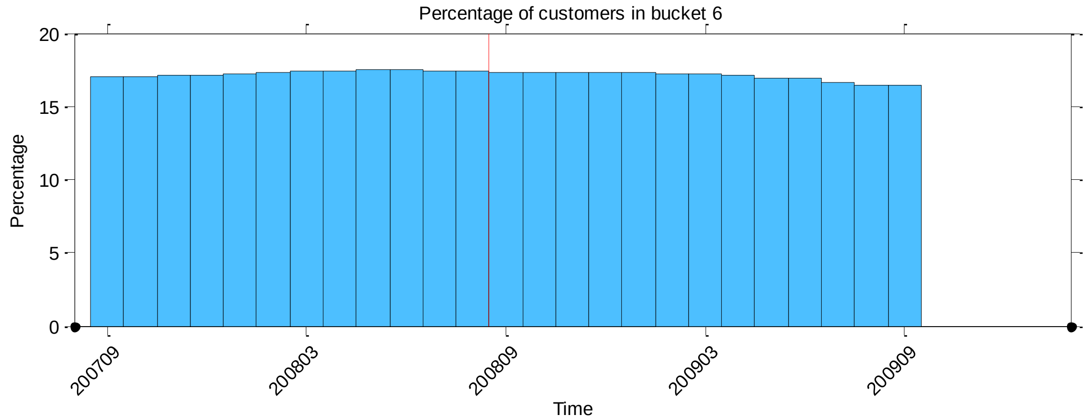


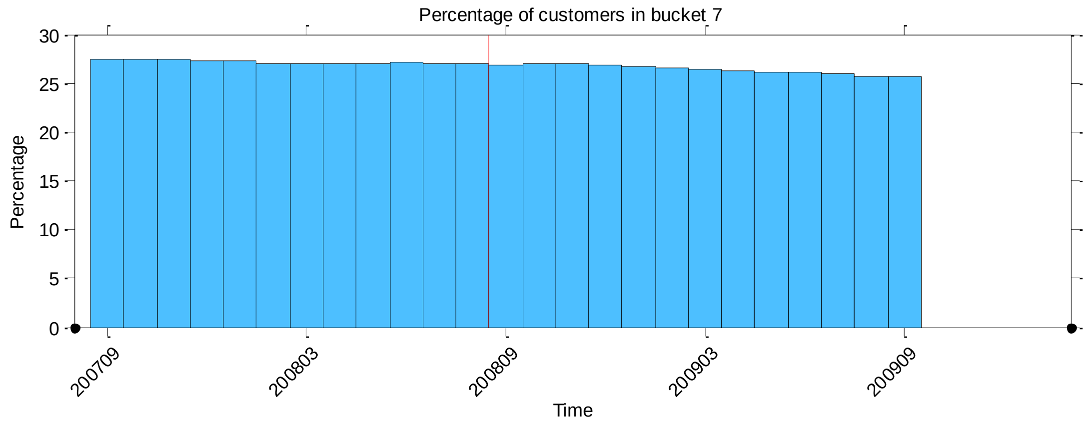


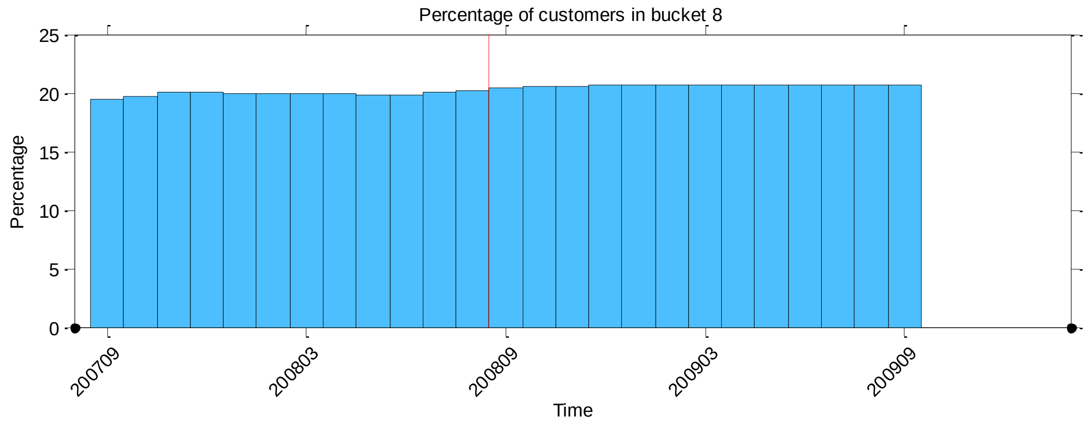

Percentage of customers in bucket 3
0.01
0.008
0.006
0.004
0.002
0
Percentage
200709
200803
200809
200903
200909
Time

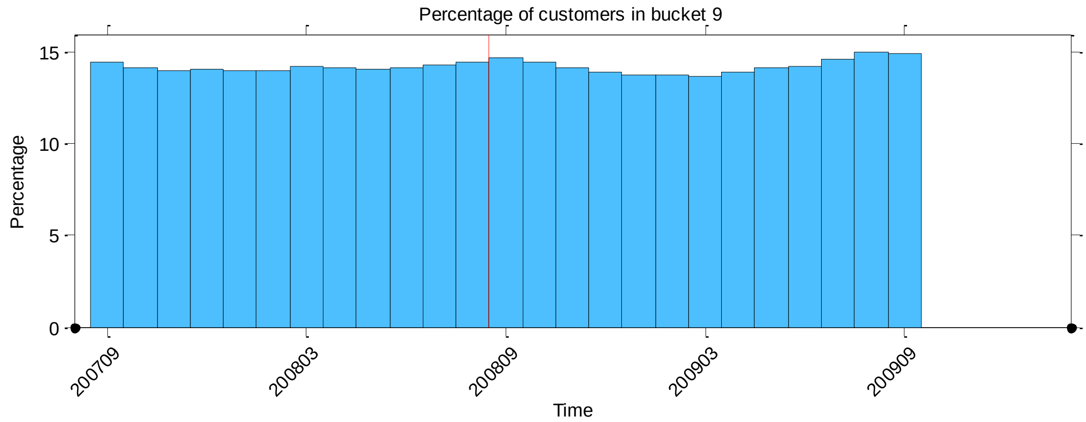

Percentage of customers in bucket 4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Percentage
200709
200803
200809
200903
200909
Time

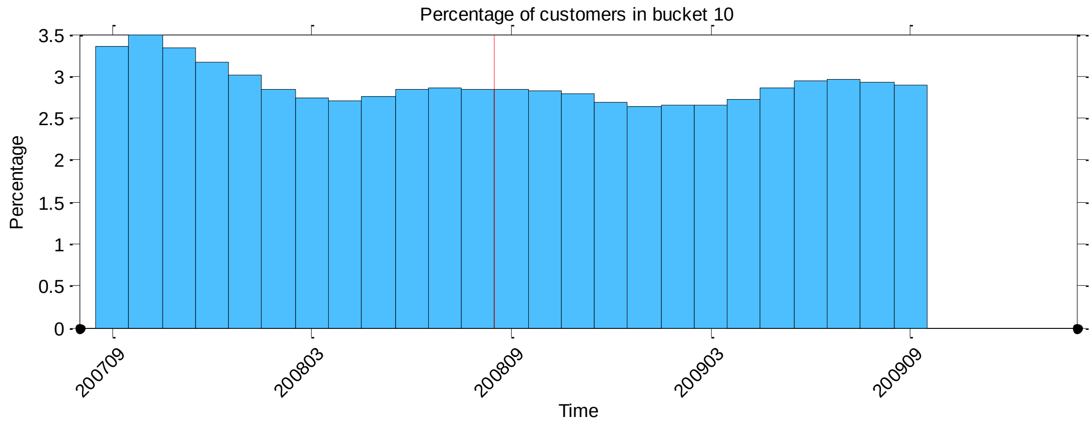

Percentage of customers in bucket 5
14
12
10
8
6
4
2
0
Percentage
200709
200803
200809
200903
200909
Time

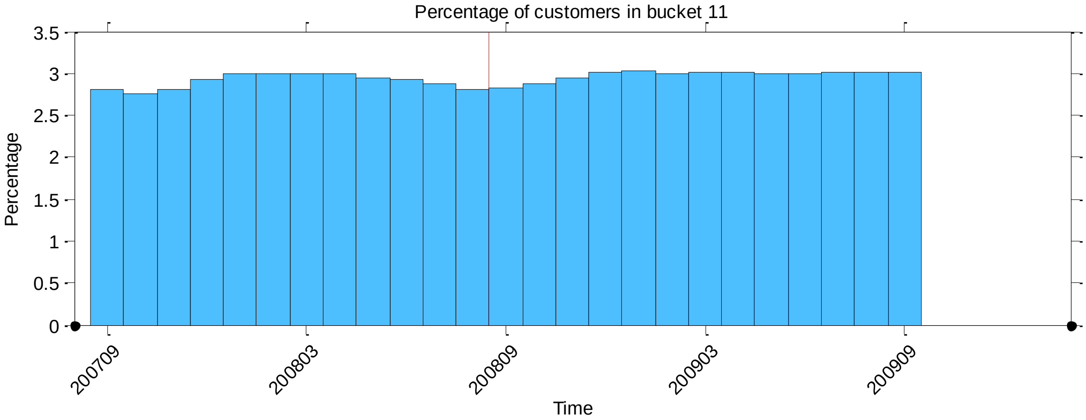

Percentage of customers in bucket 6
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

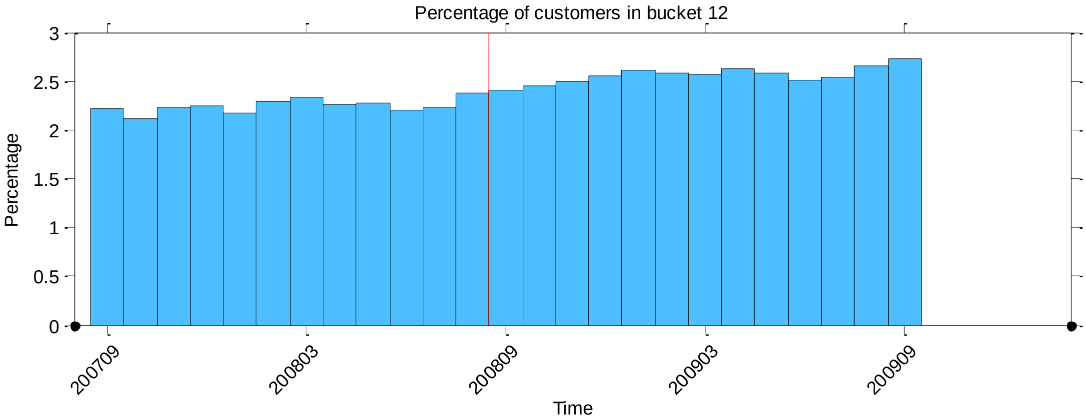

Percentage of customers in bucket 7
Percentage
30
25
20
15
10
5
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 8
Percentage
25
20
15
10
5
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 9
Percentage
15
10
5
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 10
Percentage
3.5
3
2.5
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time

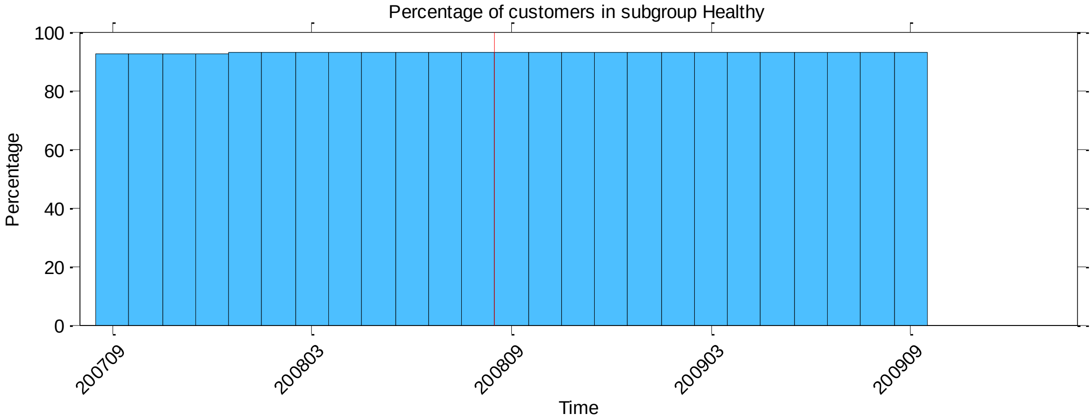

Percentage of customers in bucket 11
Percentage
3.5
3
2.5
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time

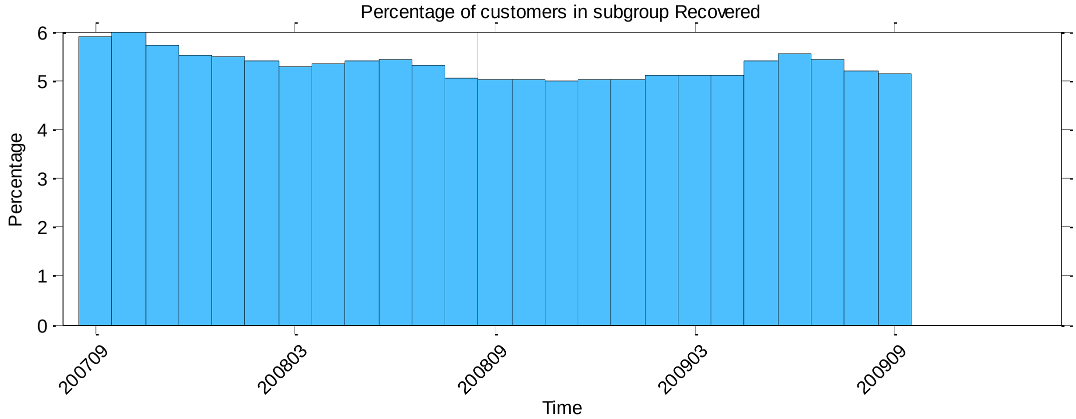

Percentage of customers in bucket 12
Percentage
3
2.5
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time

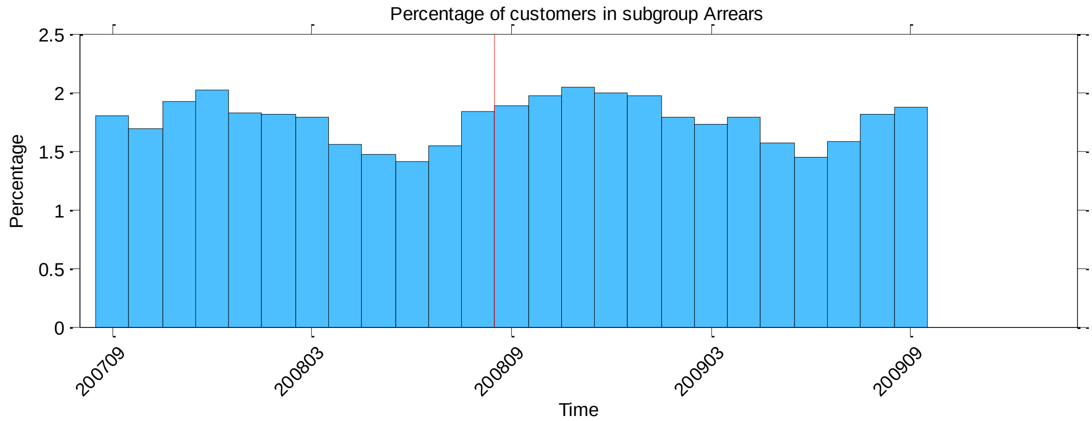
APPENDIX 13 – Logit LT prototype given 1 year PD results

This APPENDIX contains statistics and test results from the Logit LT prototype given 1 year. The mean number of observations per moment of measurement is equal to 159381.56.

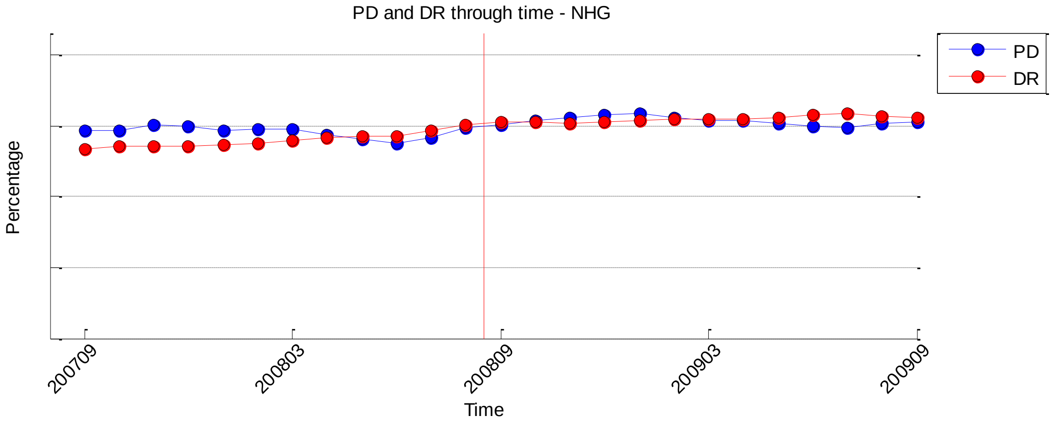
Healthy/Recovered/Default

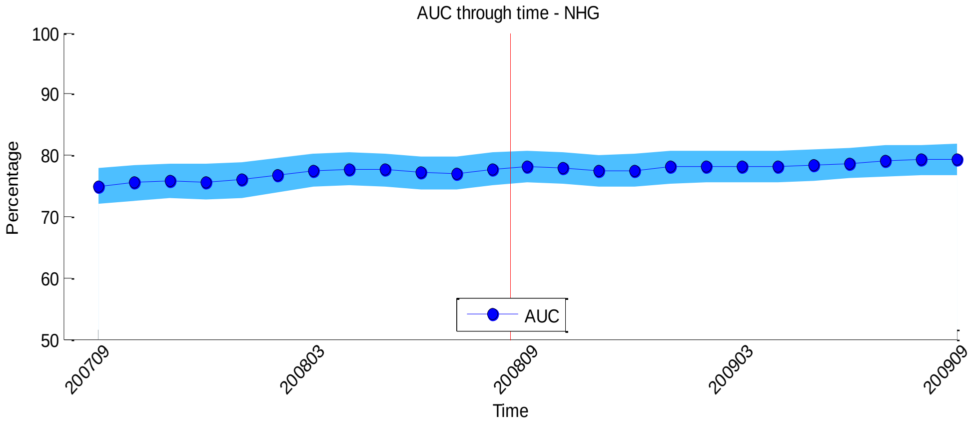


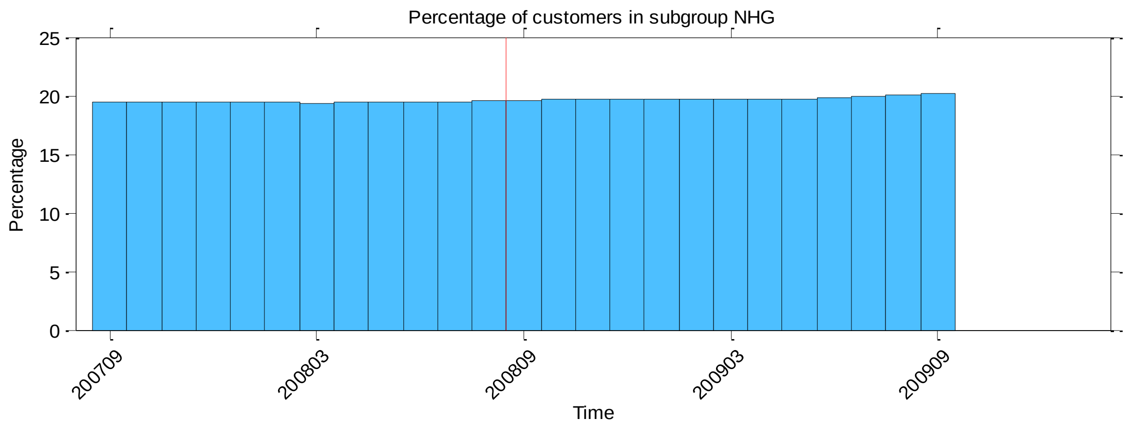


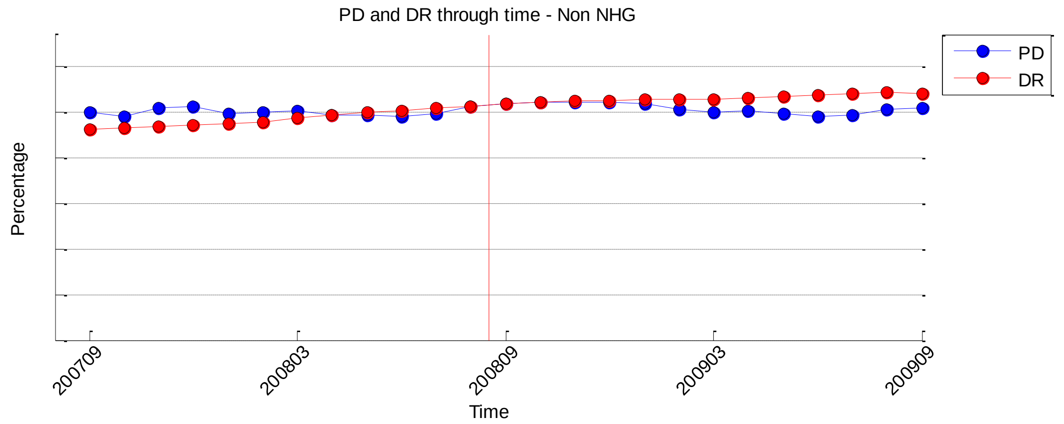


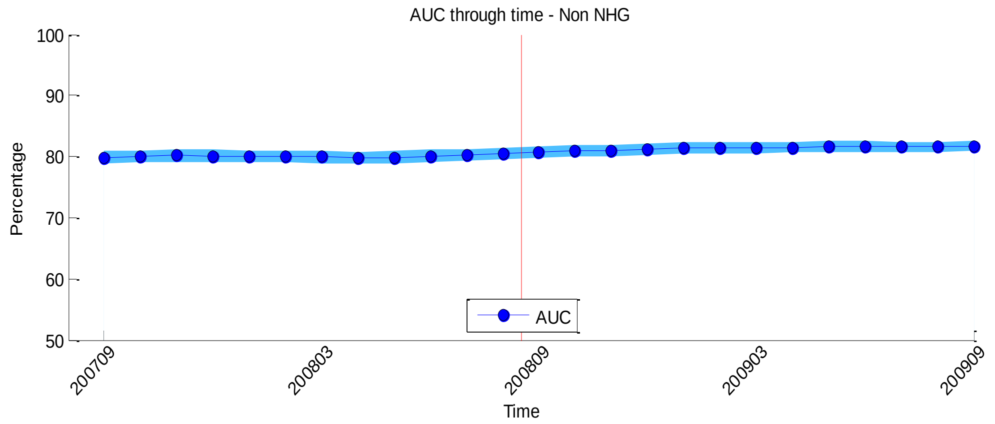
NHG/Non NHG

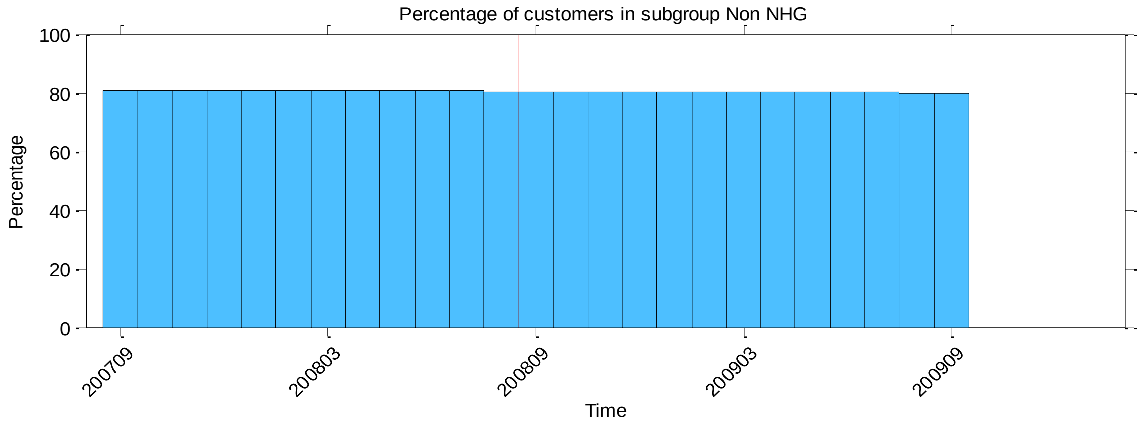

PD and DR through time - NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

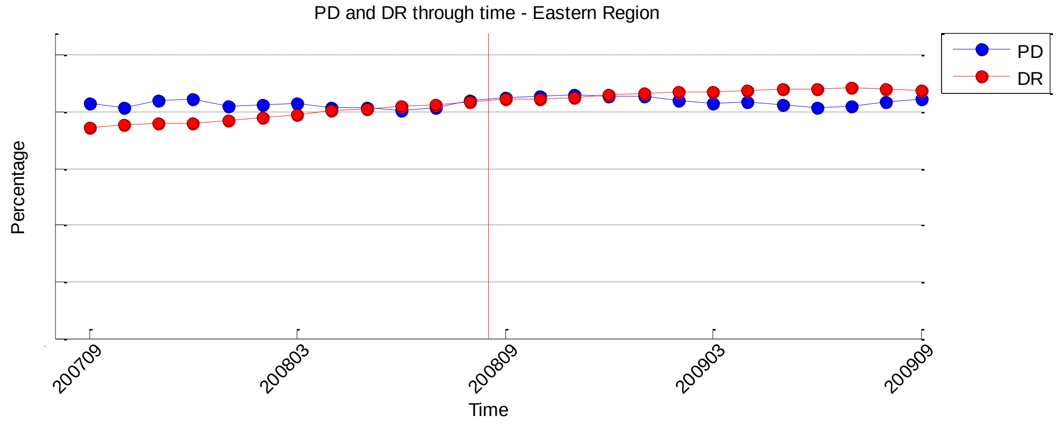

Percentage of customers in subgroup NHG
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

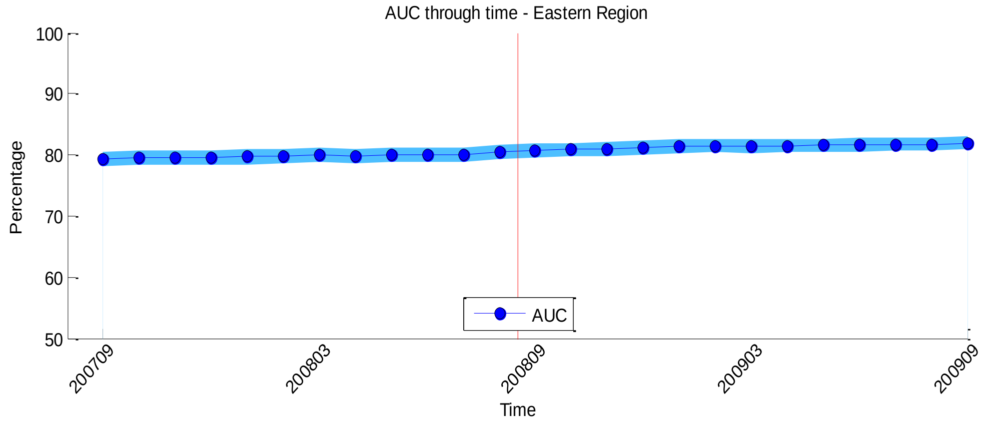

PD and DR through time - Non NHG
Percentage
PD
DR
200709
200803
200809
200903
200909
Time

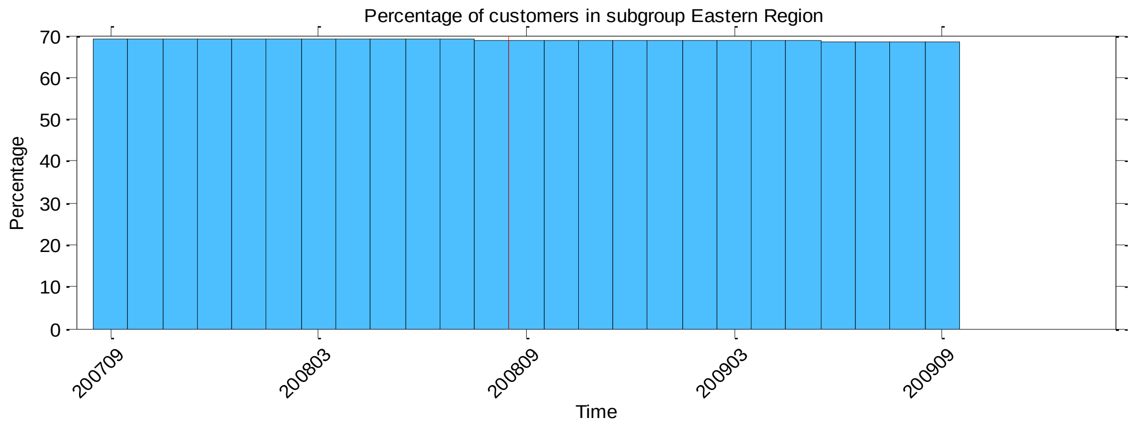

AUC through time - Non NHG
Percentage
100
90
80
70
60
50
AUC
200709
200803
200809
200903
200909
Time

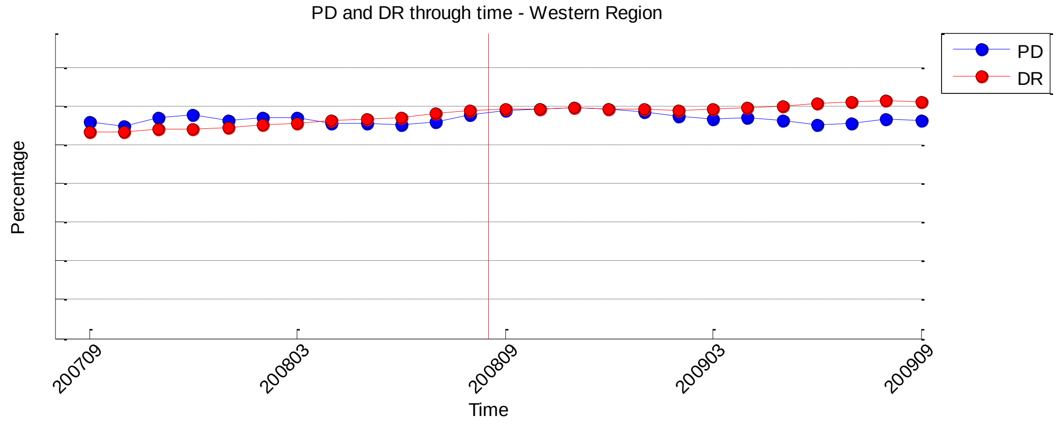

Percentage of customers in subgroup Non NHG
Percentage
100
80
60
40
20
0
200709
200803
200809
200903
200909
Time

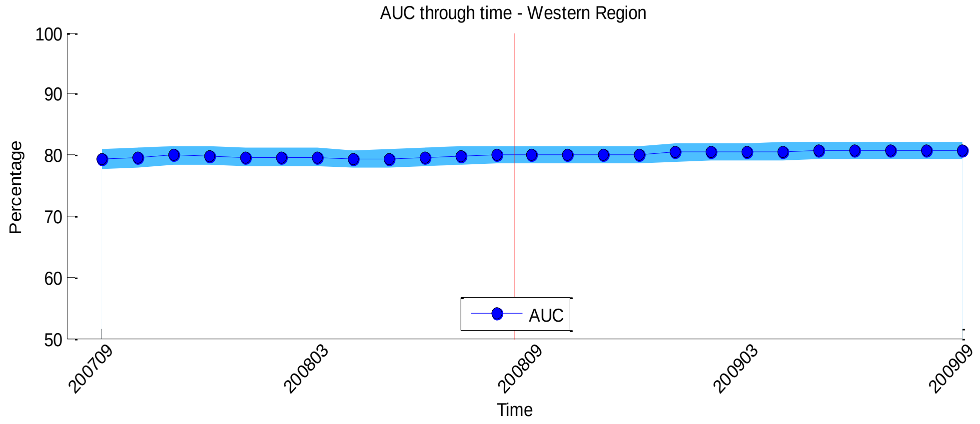
Eastern Region/Western Region

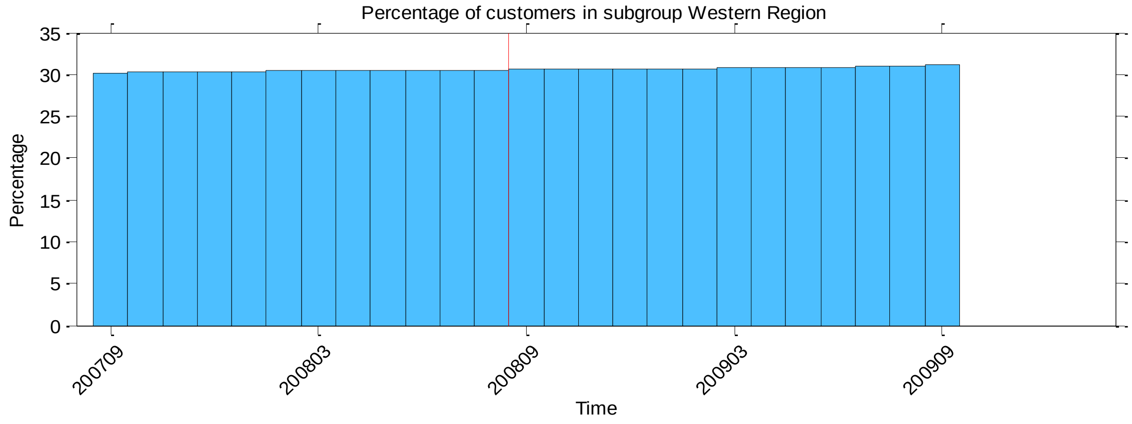






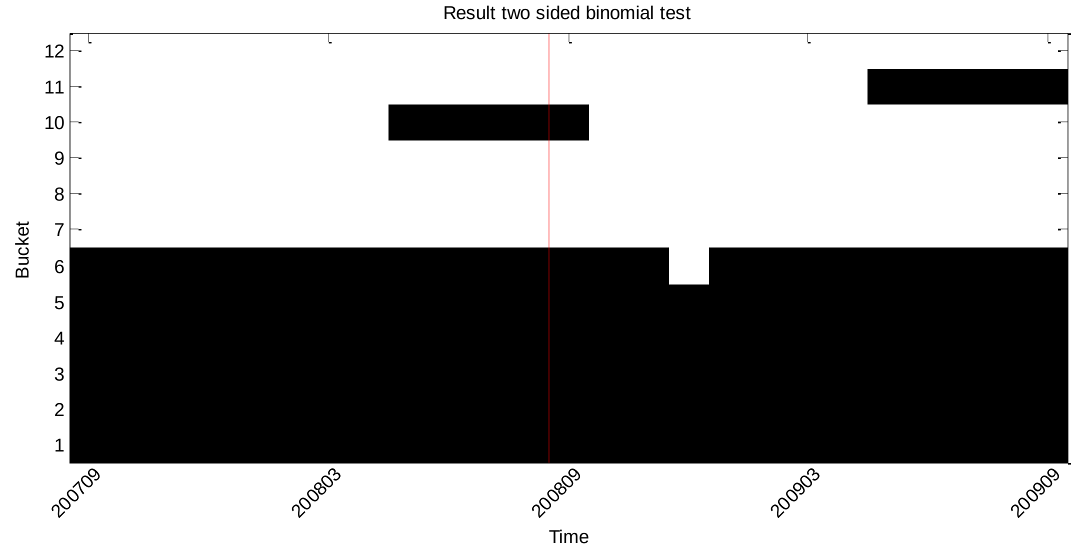

PD and DR through time - Western Region
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - Western Region
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

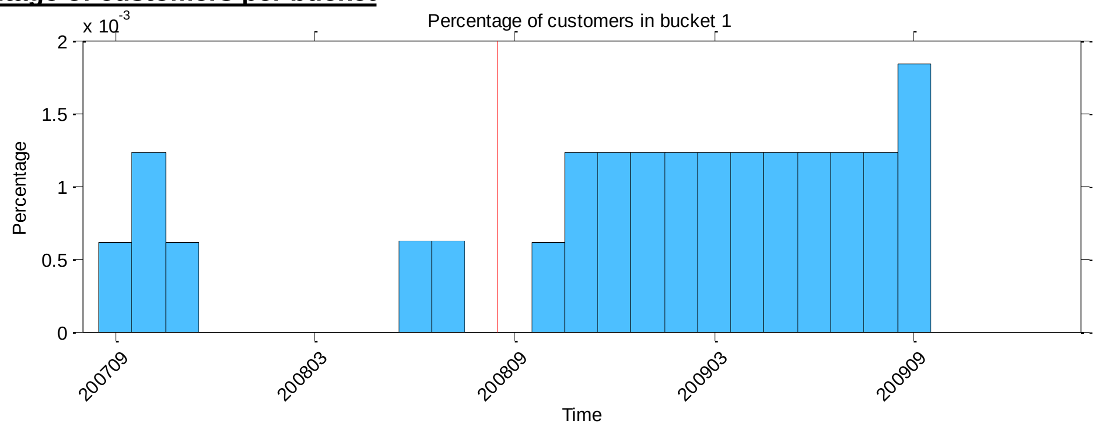

Percentage of customers in subgroup Western Region
35
30
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

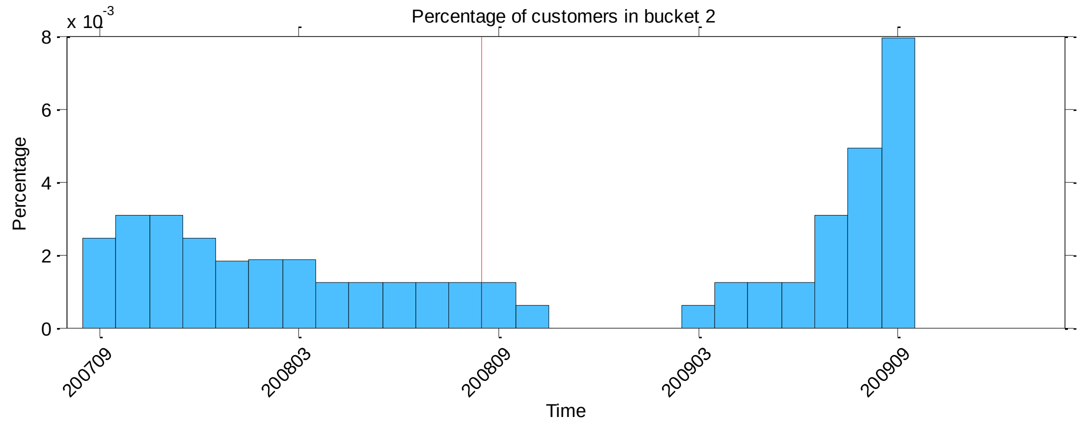
Binomial test Results on bucket level

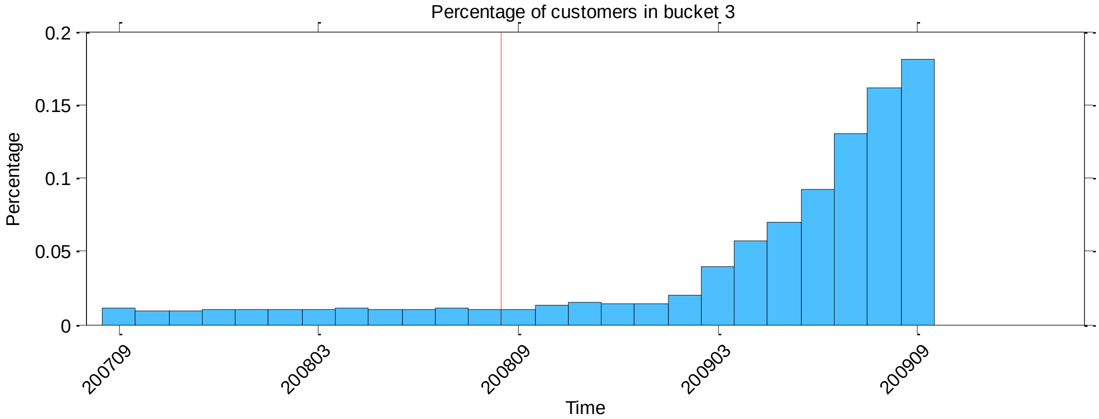
Important to notice is that the first 4 buckets contains less than 2 percent of the customers. This because a 5 year PD score is higher than a 1 year PD score. Because of the low number of observations in those buckets, the test results for those buckets cannot be seen as statistical relevant.

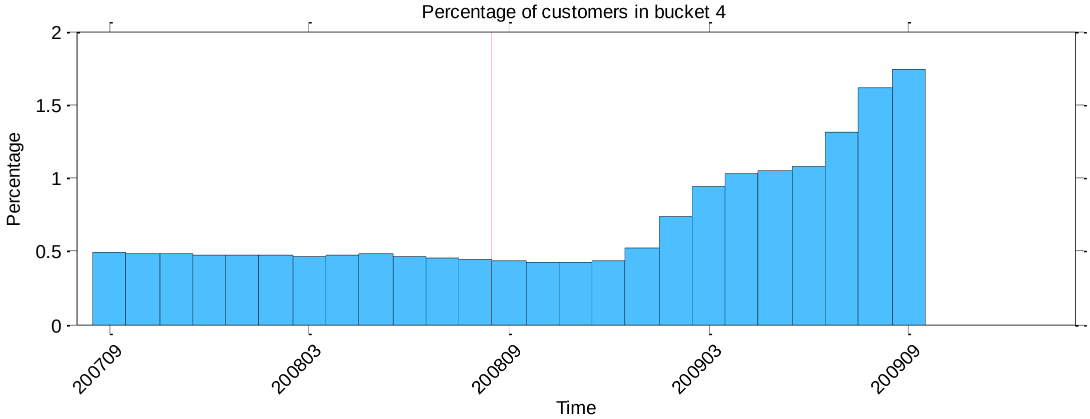


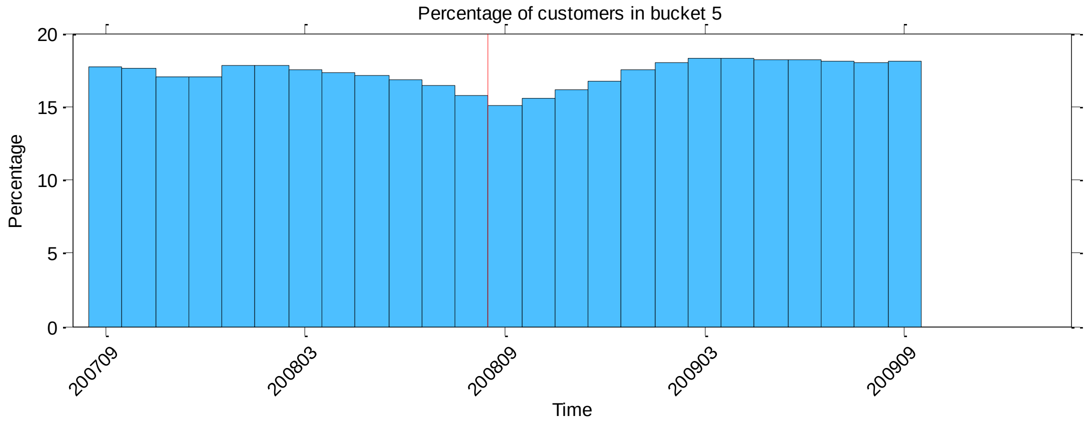
Percentage of customers per bucket

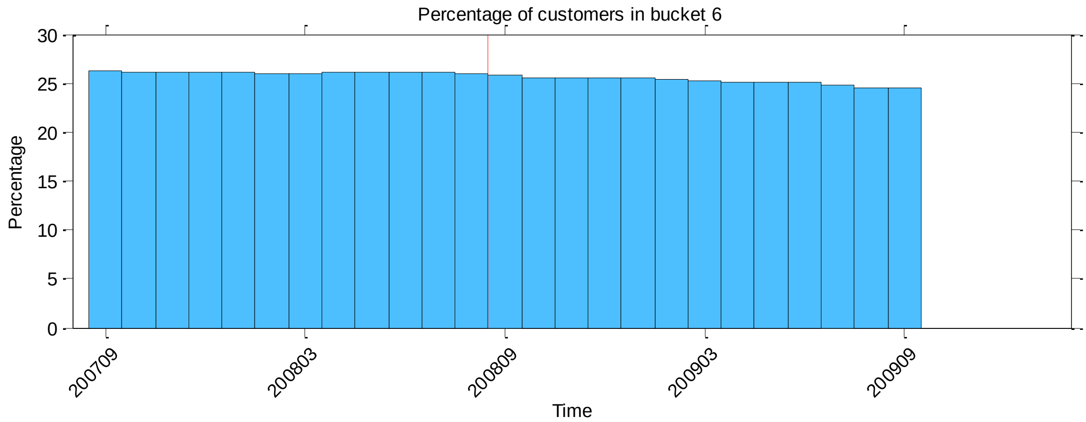


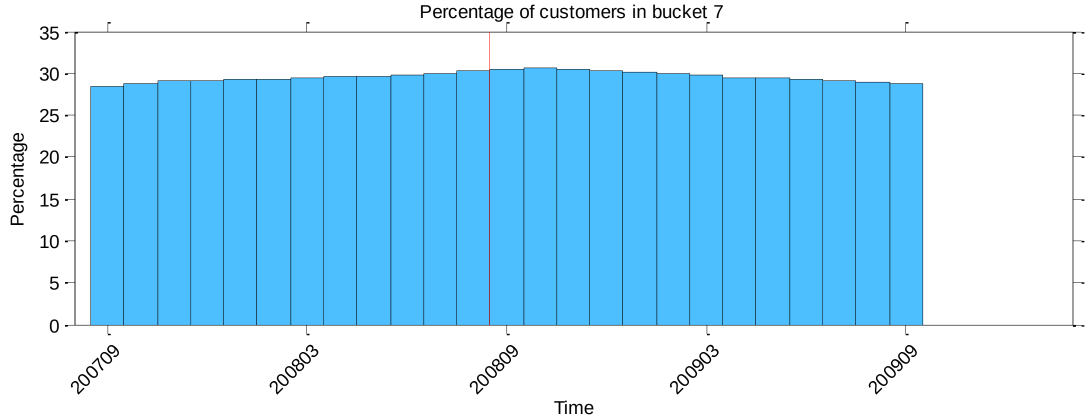



Percentage of customers in bucket 3
Percentage
0.2
0.15
0.1
0.05
0
200709
200803
200809
200903
200909
Time

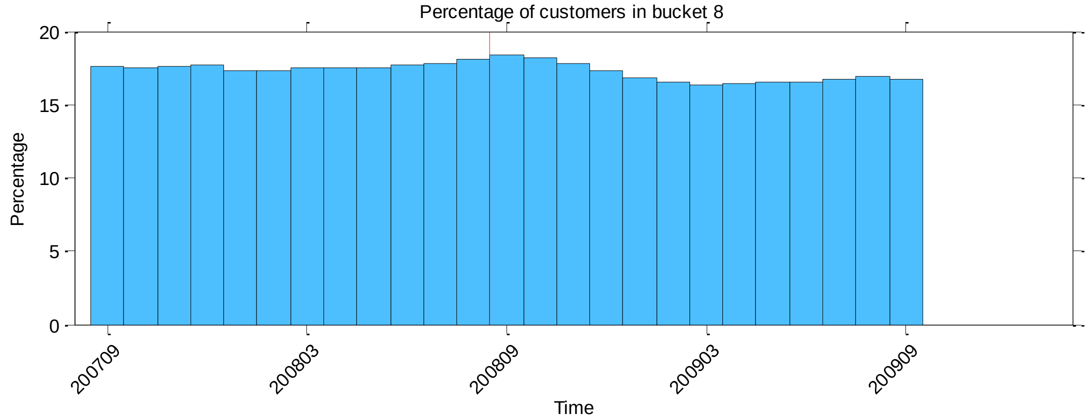

Percentage of customers in bucket 4
Percentage
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 5
Percentage
20
15
10
5
0
200709
200803
200809
200903
200909
Time

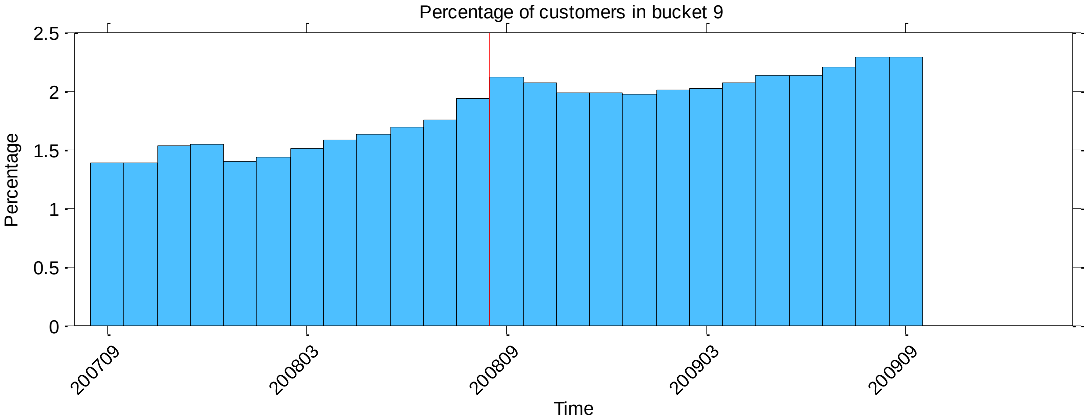

Percentage of customers in bucket 6
Percentage
30
25
20
15
10
5
0
200709
200803
200809
200903
200909
Time



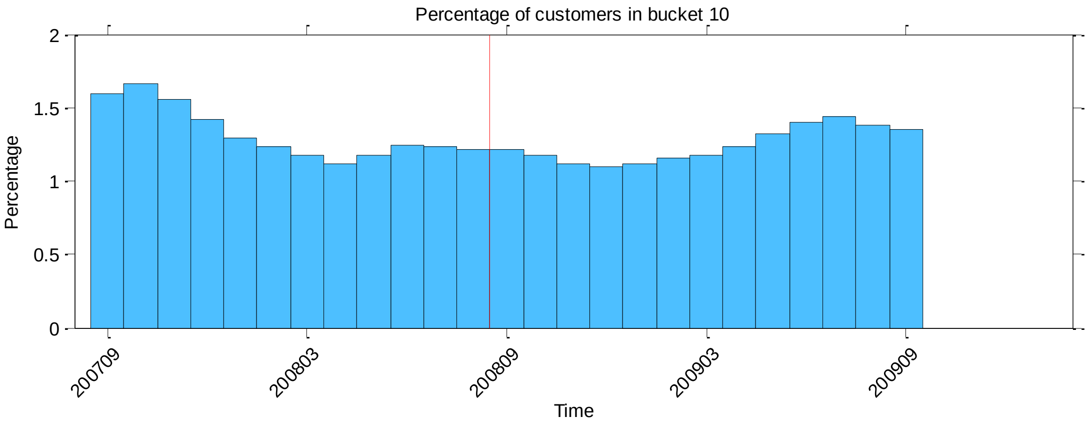
Percentage of customers in bucket 7
Percentage
Time
200709
200803
200809
200903
200909



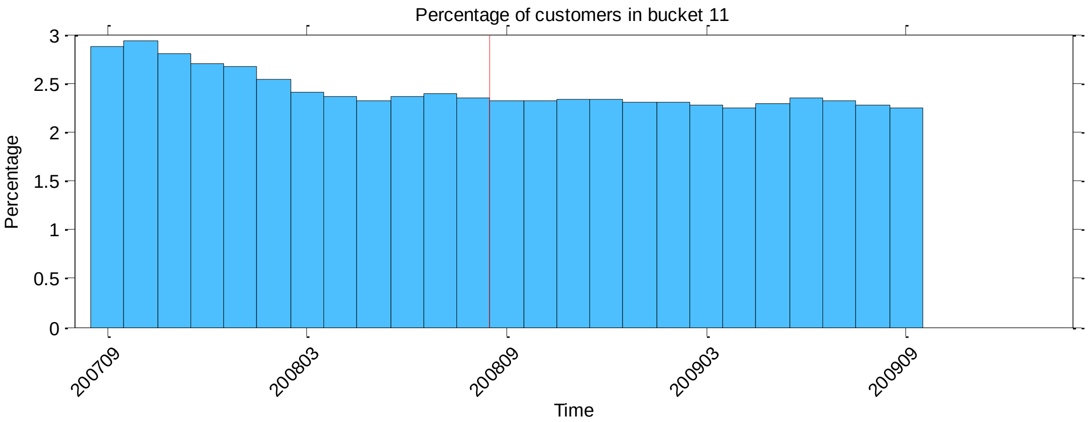
Percentage of customers in bucket 8
Percentage
Time
200709
200803
200809
200903
200909

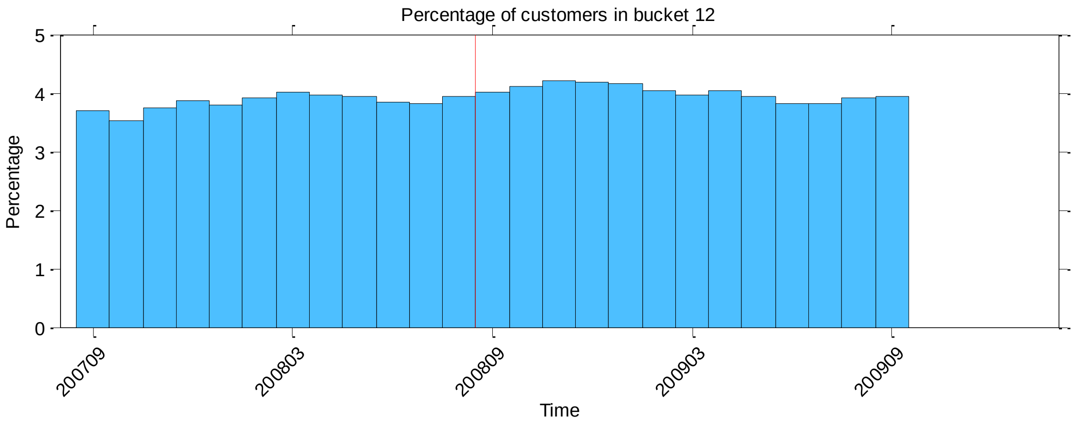


Percentage of customers in bucket 9
Percentage
Time
200709
200803
200809
200903
200909



Percentage of customers in bucket 10
Percentage
Time
200709
200803
200809
200903
200909


Percentage of customers in bucket 11
Percentage
Time
200709
200803
200809
200903
200909

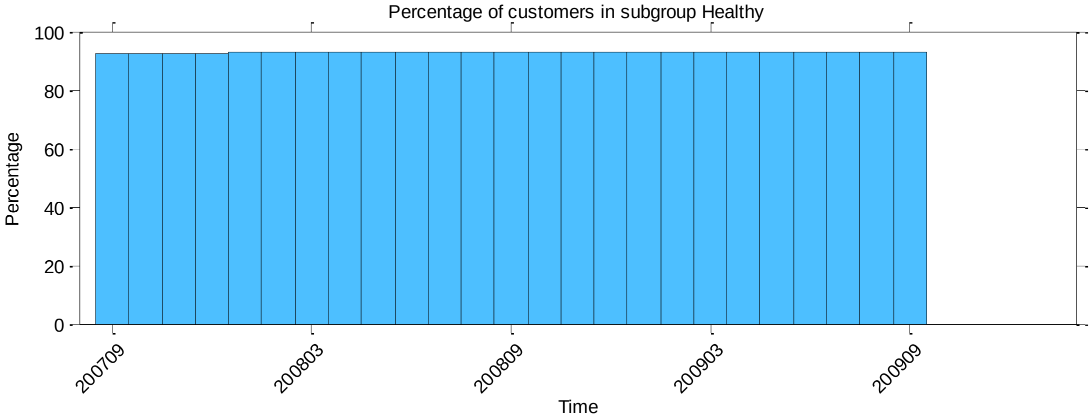

Percentage of customers in bucket 12
Percentage
Time
200709
200803
200809
200903
200909

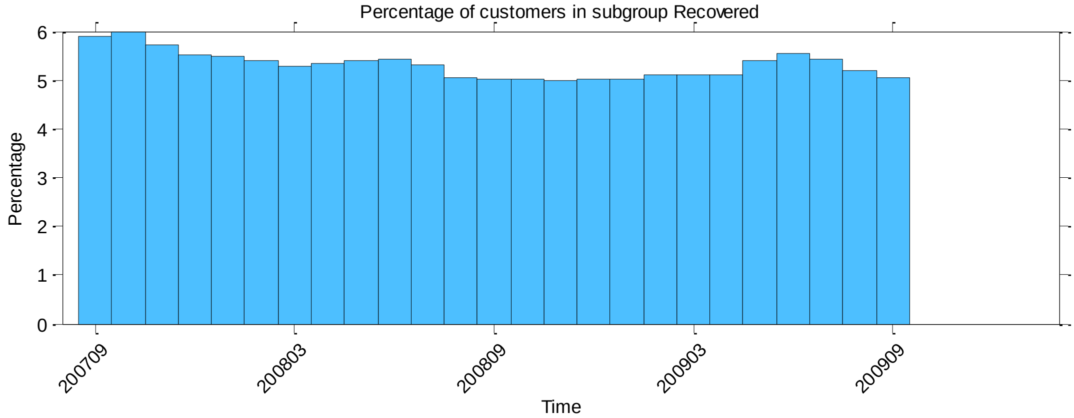
APPENDIX 14 – Hazard prototype 1 year results

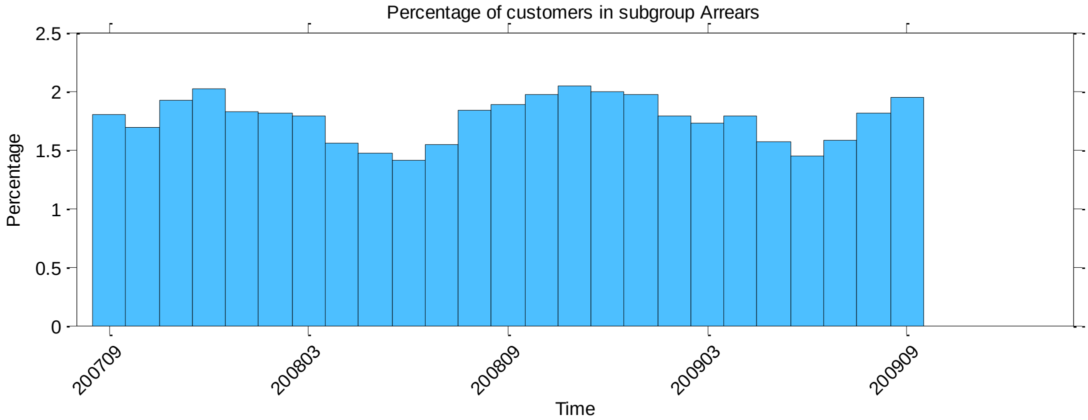
This APPENDIX contains the 1 year statistics and test results from the Hazard prototype. The mean number of observations per moment of measurement is equal to 161050.4247.

Healthy/Recovered/Default

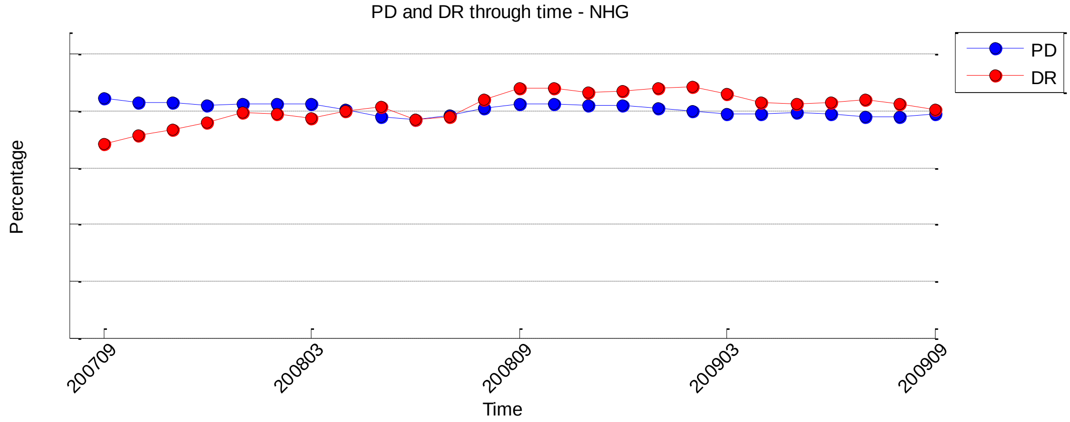
The figures below illustrates the percentage of customers per section at each moment of measurement in the observation period. This should not be confused with the percentage of customers per section at moment of estimation.

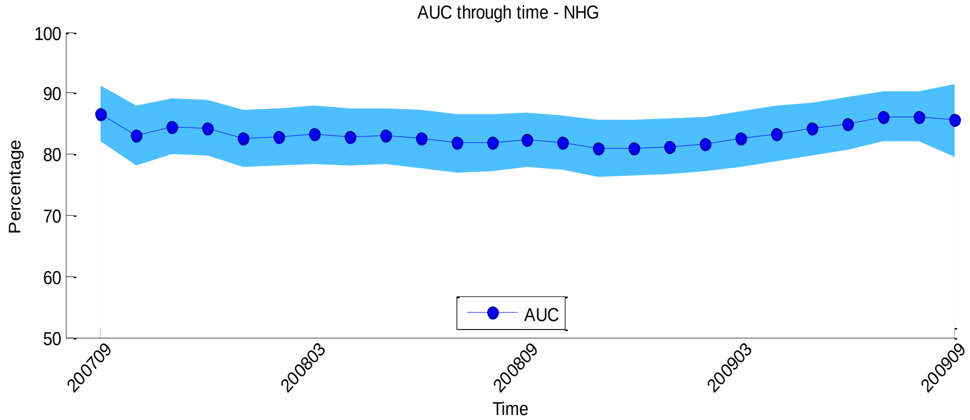


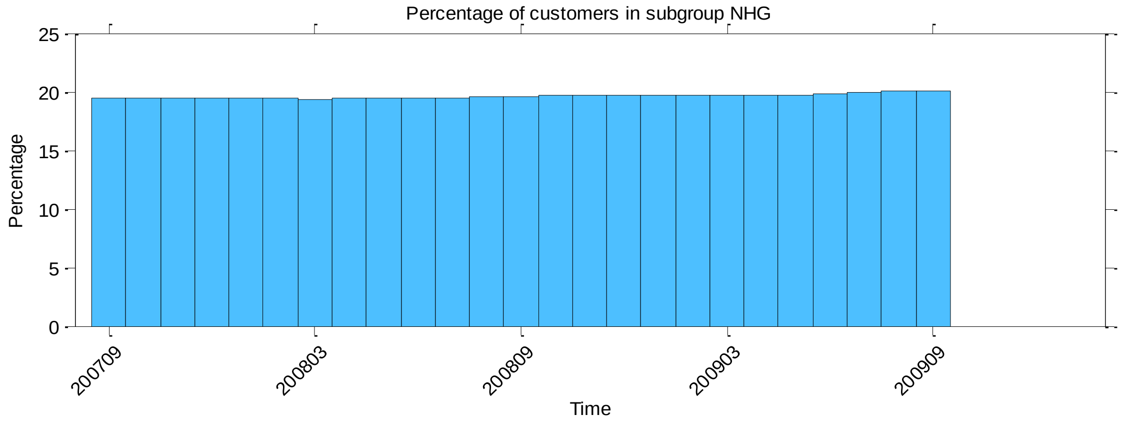


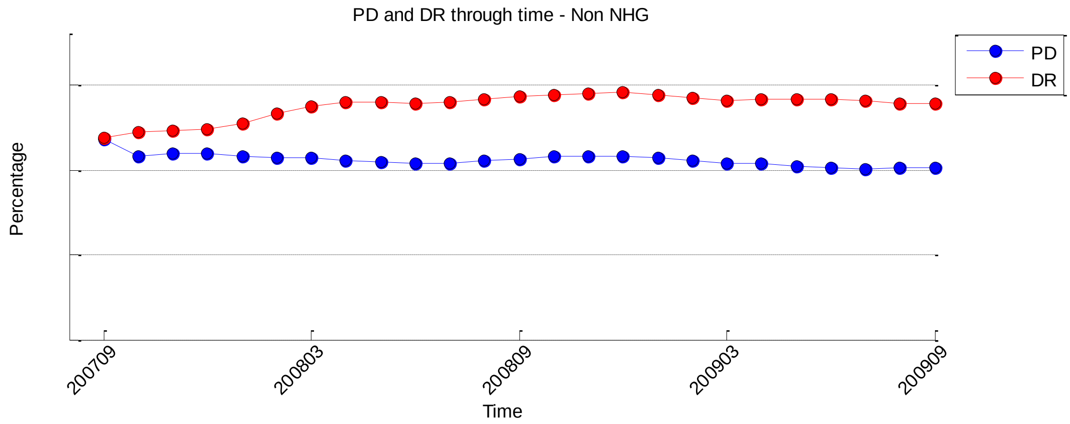


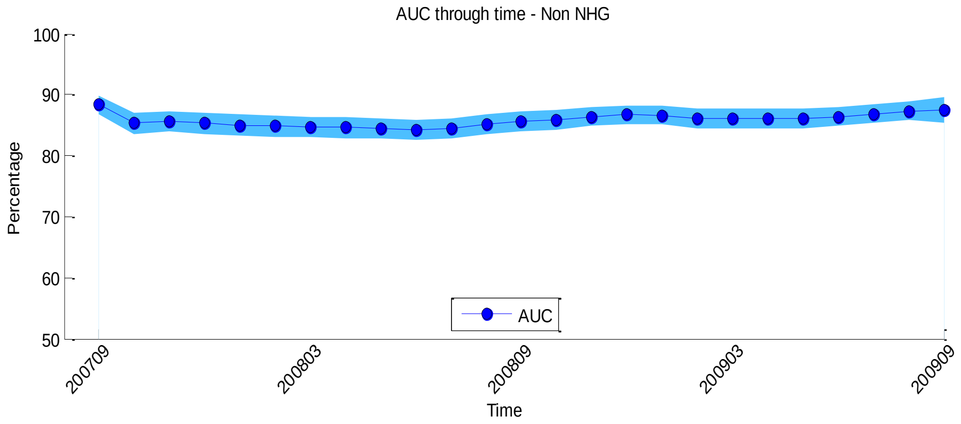
NHG/Non NHG

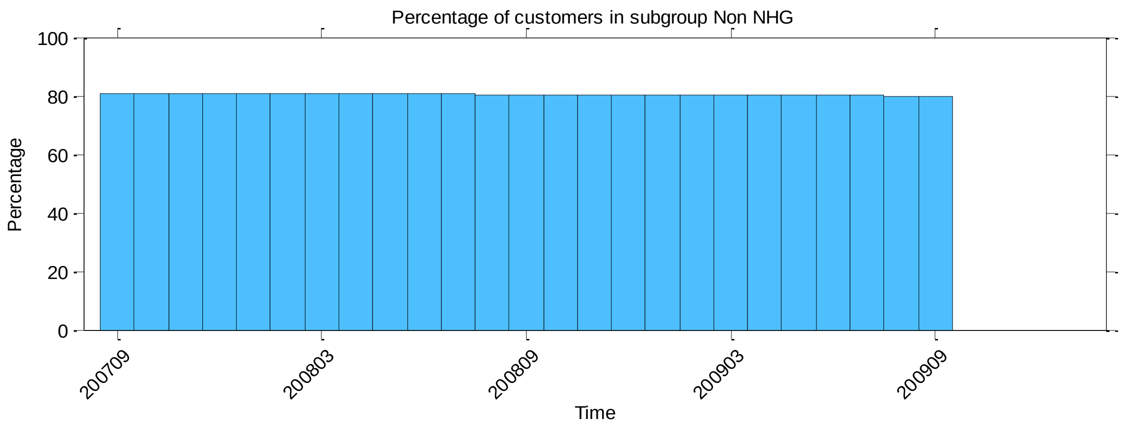

PD and DR through time - NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

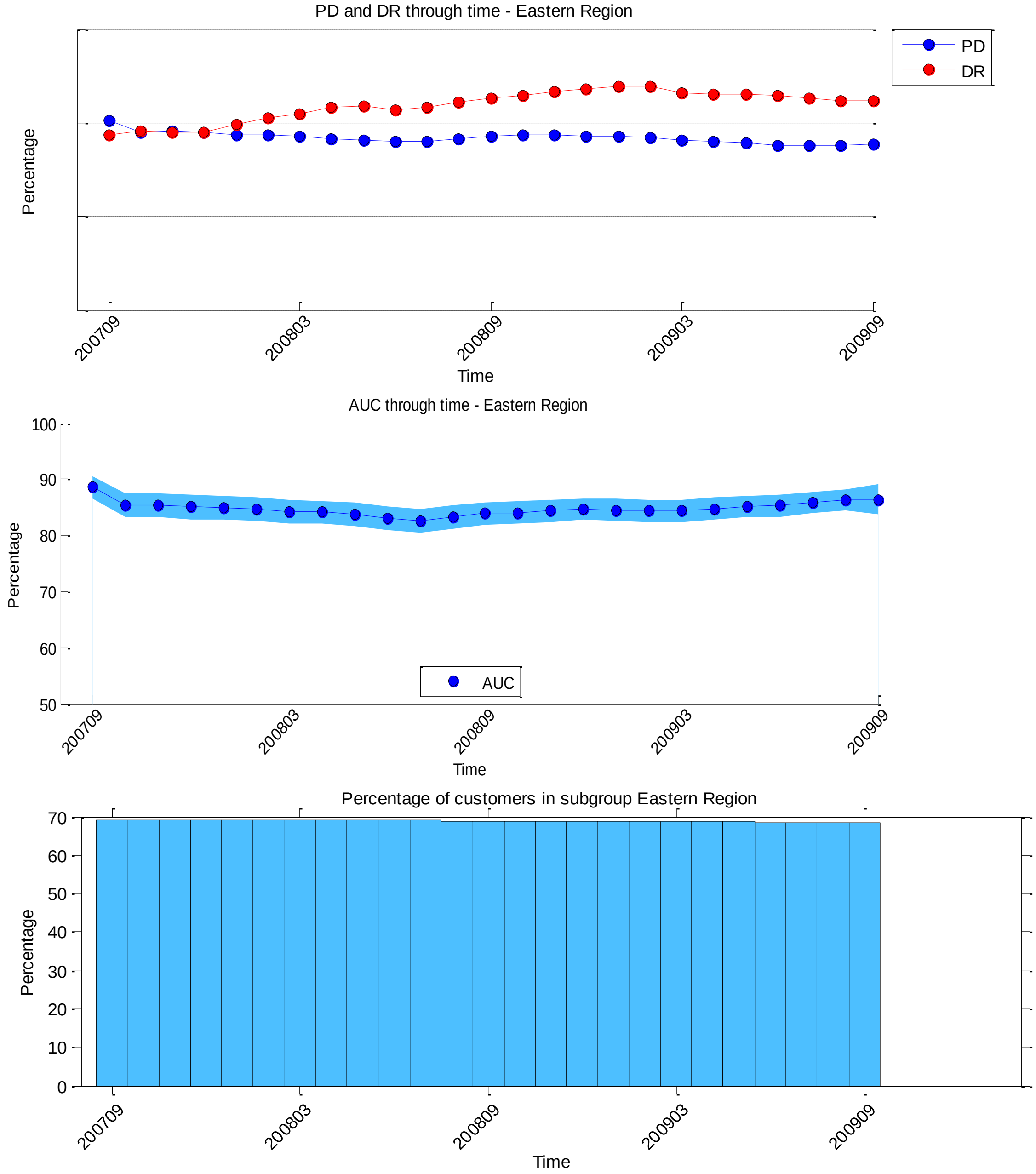

Percentage of customers in subgroup NHG
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

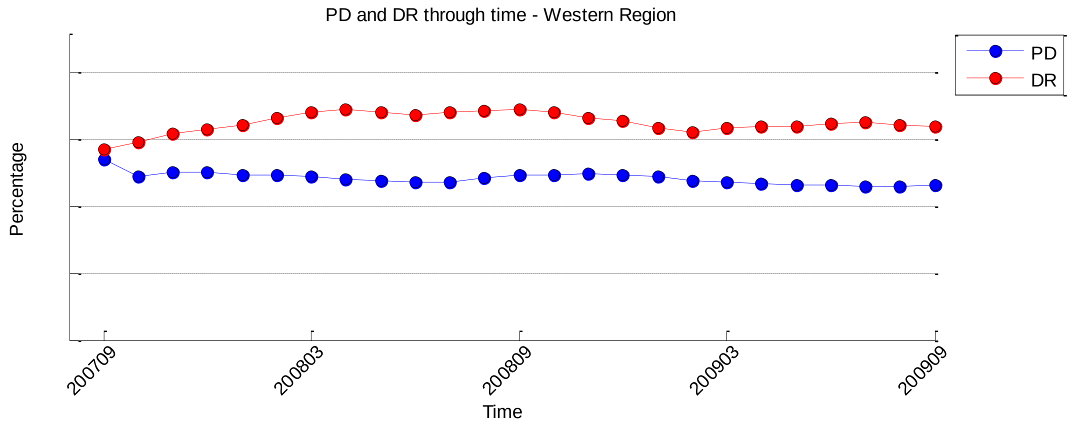

PD and DR through time - Non NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time

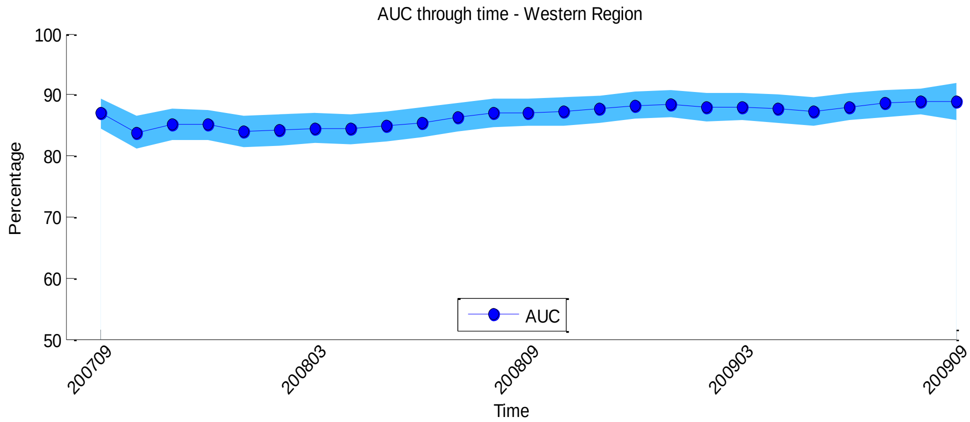

AUC through time - Non NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

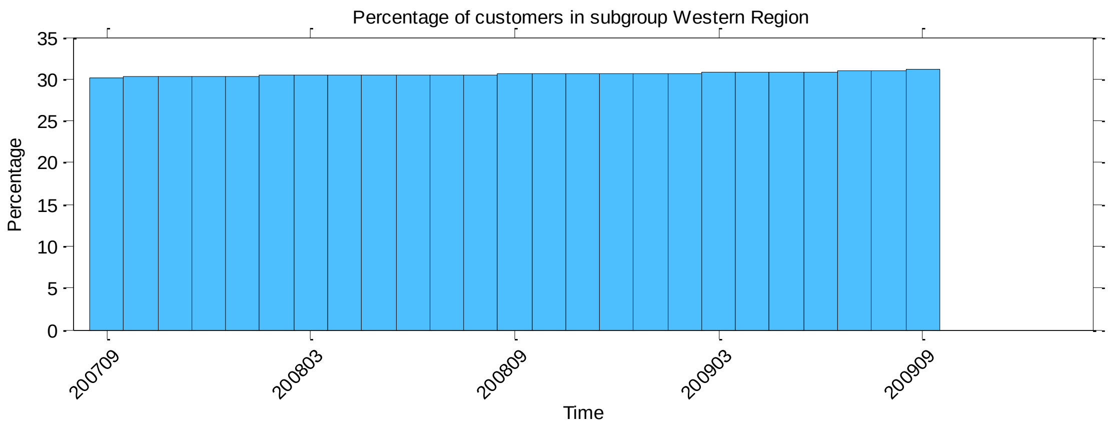

Percentage of customers in subgroup Non NHG
100
80
60
40
20
0
Percentage
200709
200803
200809
200903
200909
Time

Eastern Region/Western Region



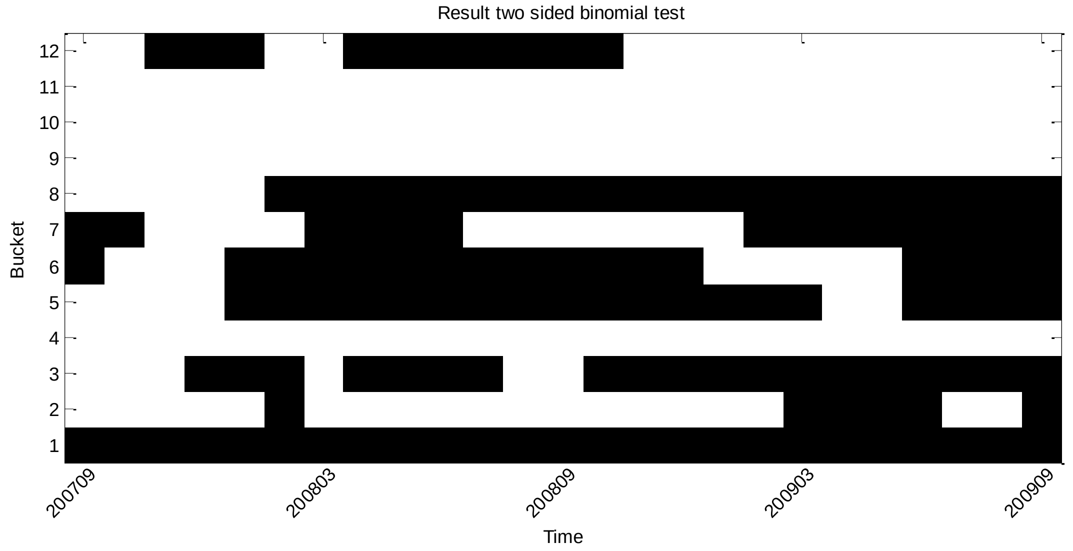

PD and DR through time - Western Region
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - Western Region
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

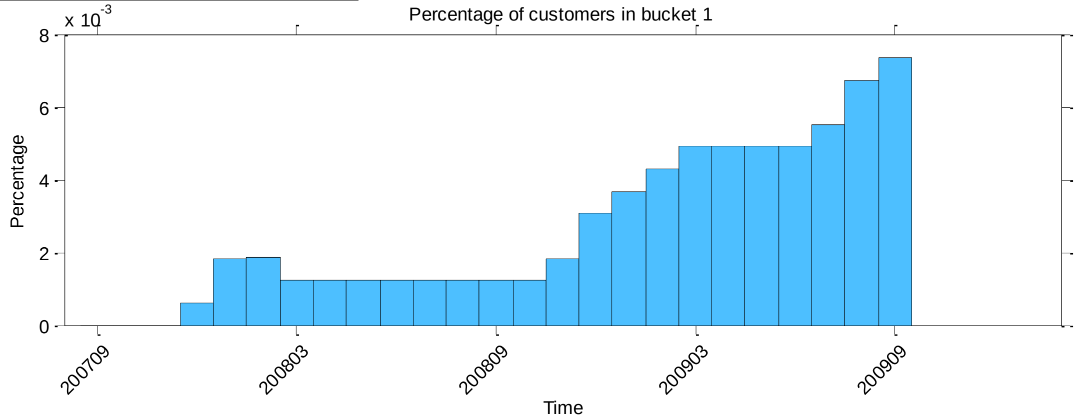

Percentage of customers in subgroup Western Region
35
30
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

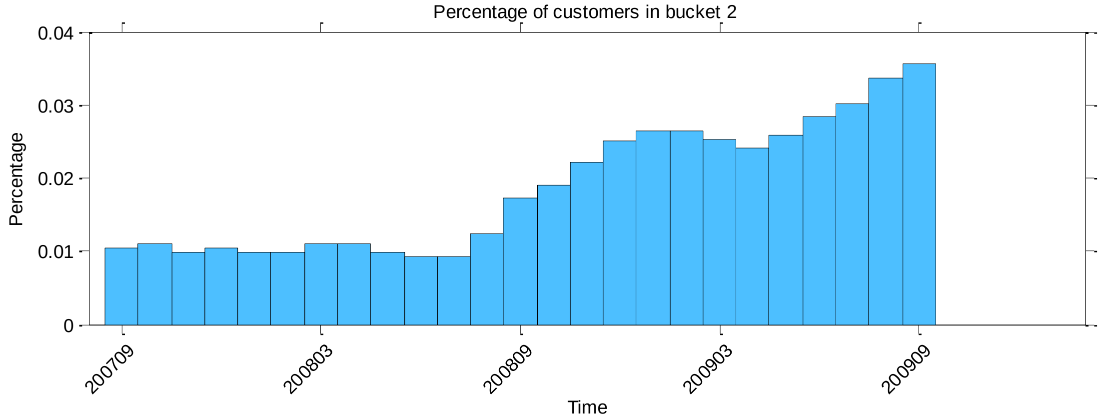
Binomial test Results on bucket level

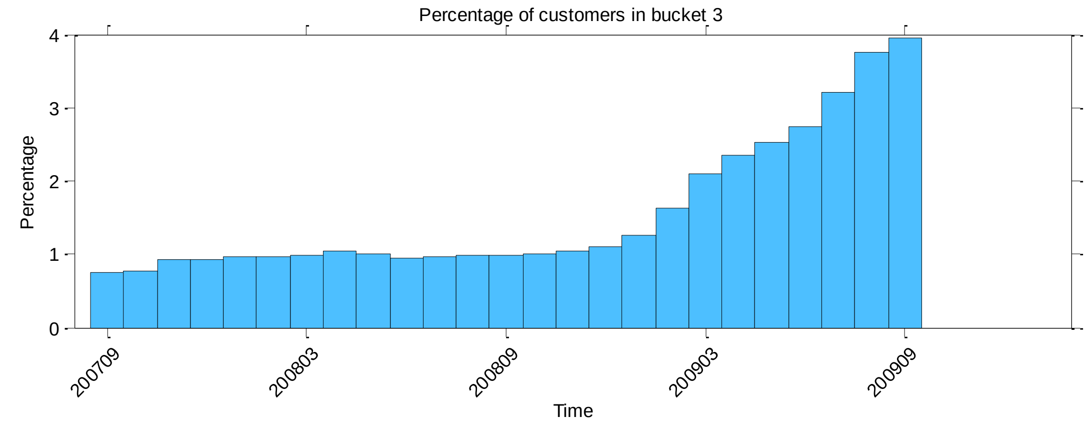
Important to notice is that the first buckets 1, 2, 7 and 12 contains together less than 2 percent of the customers. Because of the low number of observations in those buckets, the test results for those buckets cannot be seen as statistical relevant.

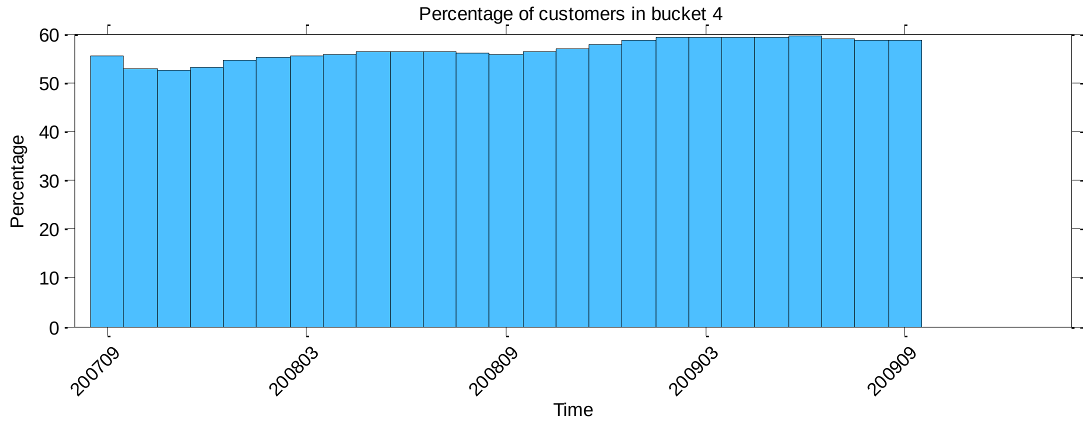


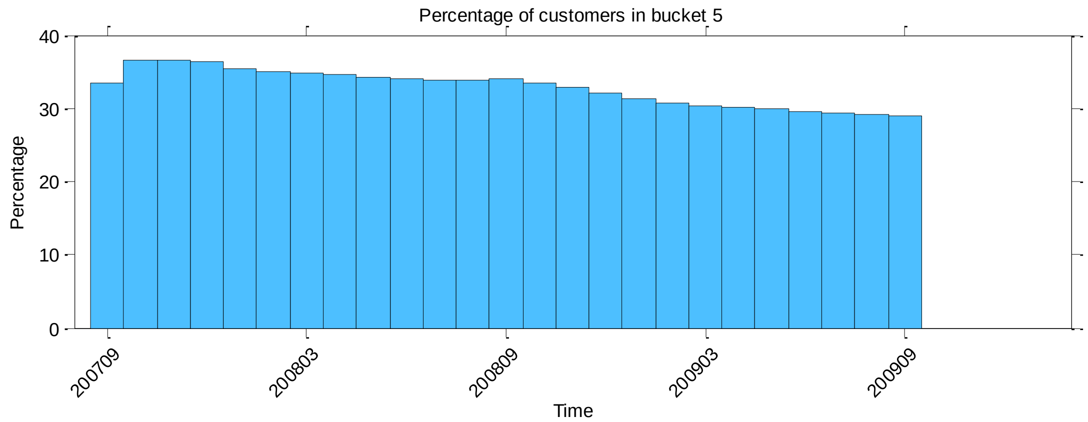
Percentage of customers per bucket

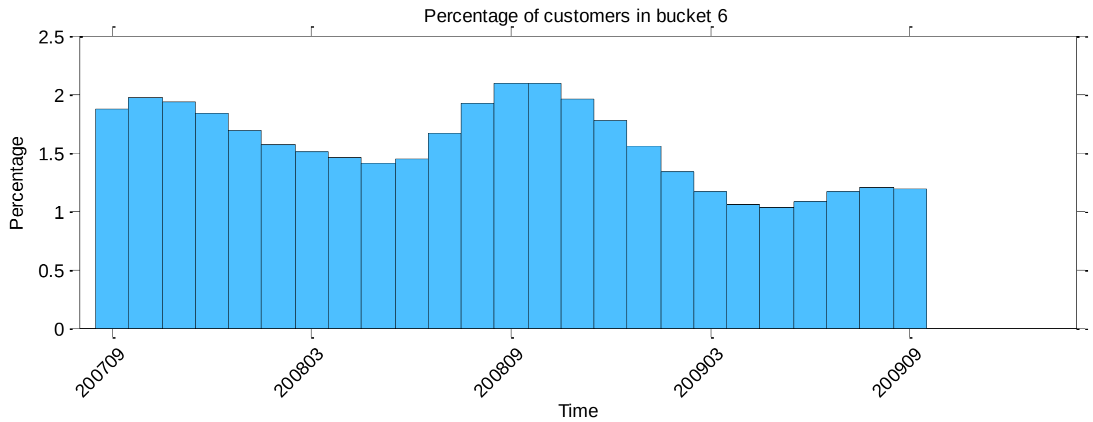


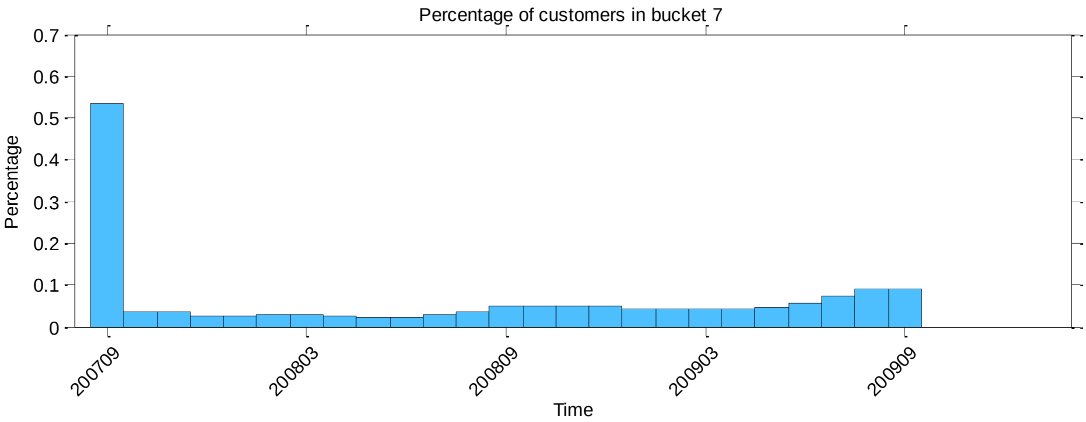


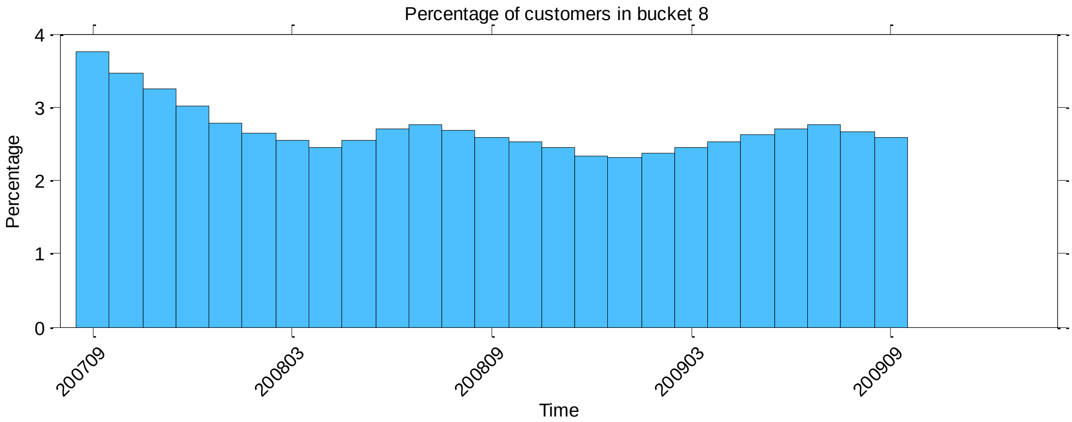

Percentage of customers in bucket 3
Percentage
Time
200709
200803
200809
200903
200909

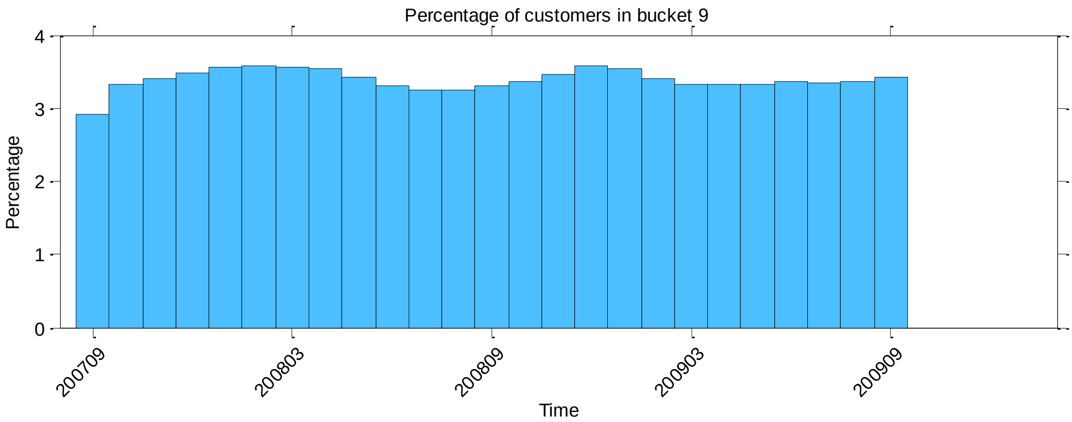

Percentage of customers in bucket 4
Percentage
Time
200709
200803
200809
200903
200909

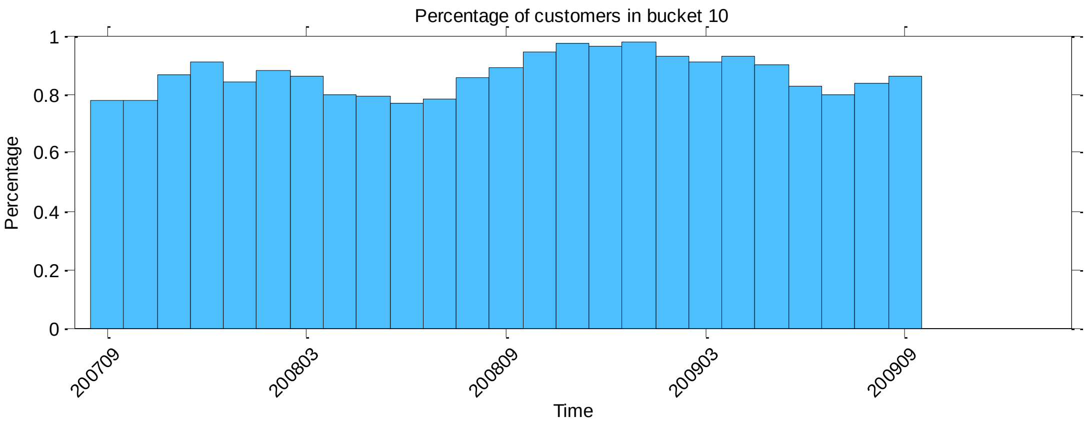

Percentage of customers in bucket 5
Percentage
Time
200709
200803
200809
200903
200909

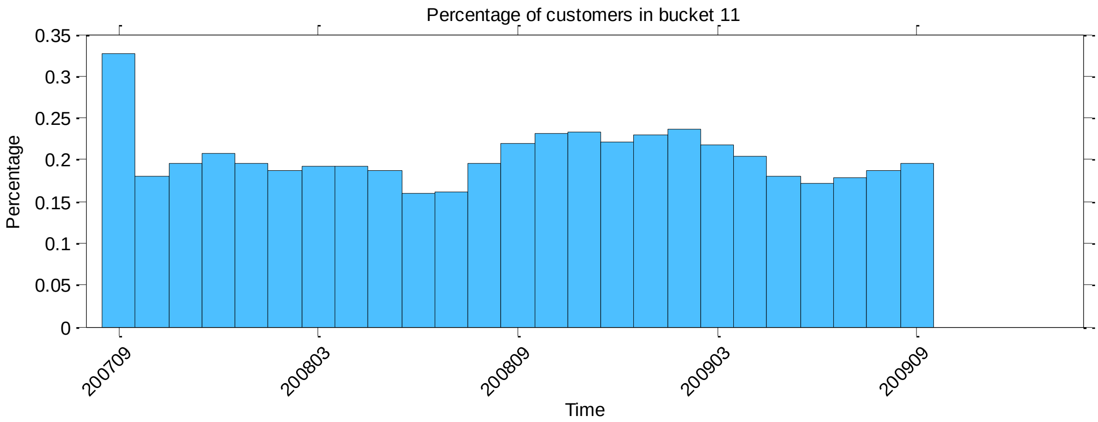

Percentage of customers in bucket 6
Percentage
Time
200709
200803
200809
200903
200909

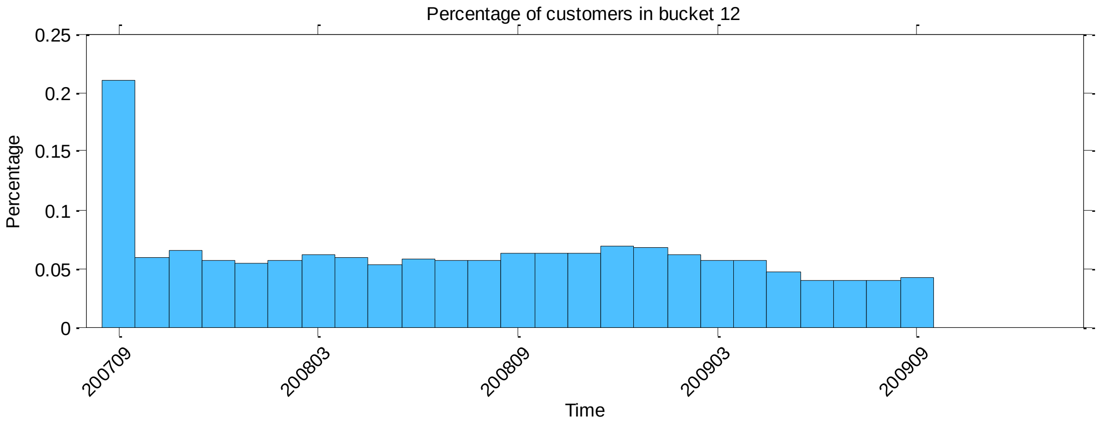

Percentage of customers in bucket 7
Percentage
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 8
Percentage
4
3
2
1
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 9
Percentage
4
3
2
1
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 10
Percentage
1
0.8
0.6
0.4
0.2
0
200709
200803
200809
200903
200909
Time

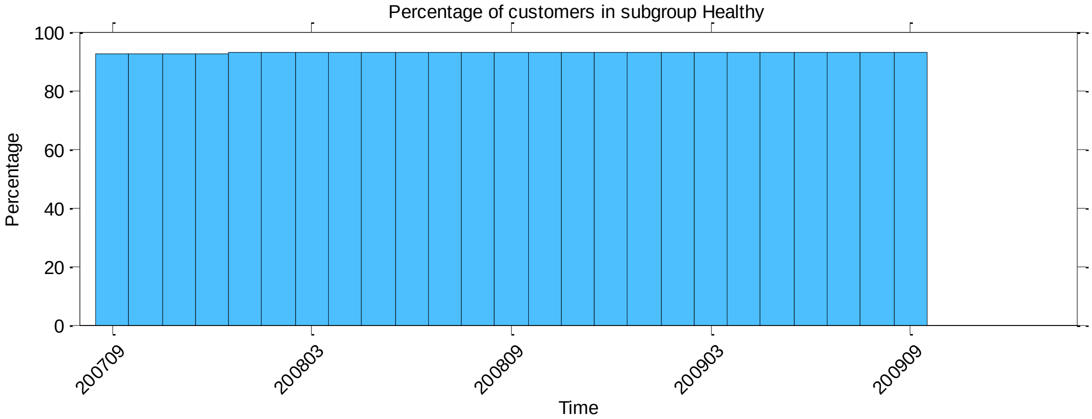

Percentage of customers in bucket 11
Percentage
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
200709
200803
200809
200903
200909
Time

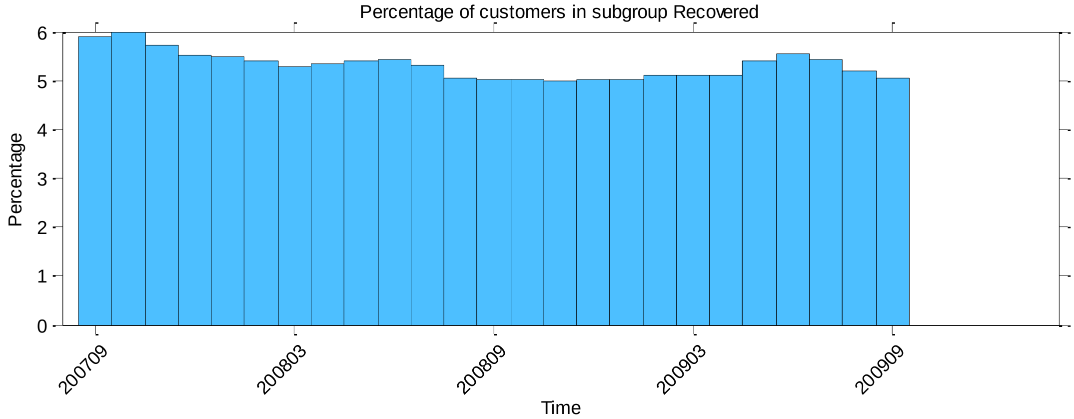

Percentage of customers in bucket 12
Percentage
0.25
0.2
0.15
0.1
0.05
0
200709
200803
200809
200903
200909
Time

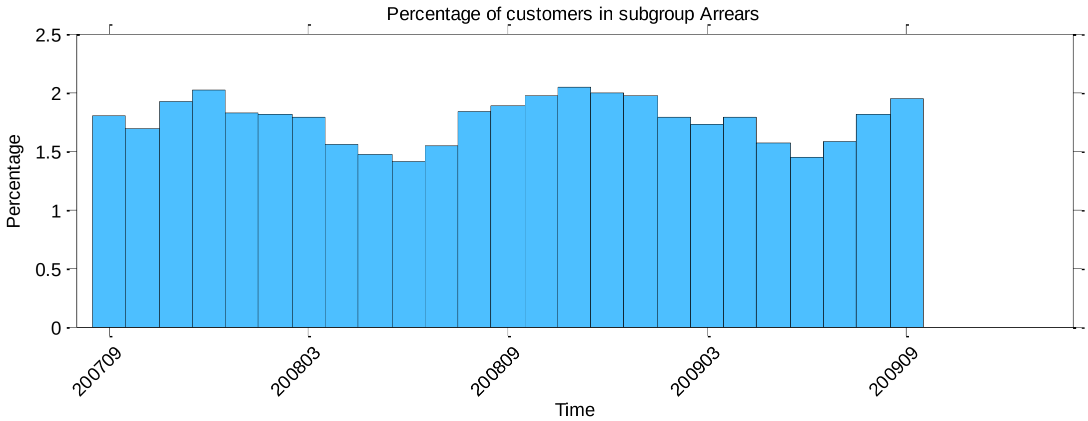
APPENDIX 15 – Hazard prototype LT results

This APPENDIX contains the 5 year statistics and test results from the Hazard prototype. The mean number of observations per moment of measurement is equal to 159381.56.

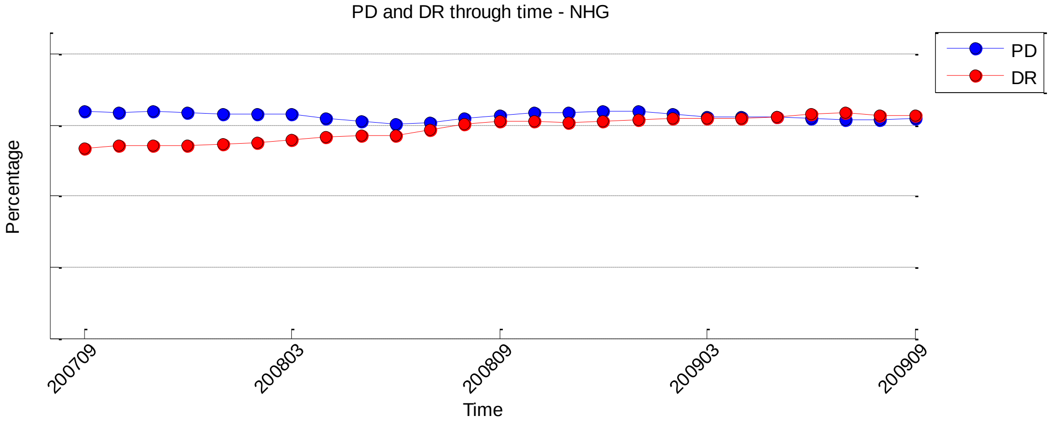
Healthy/Recovered/Default

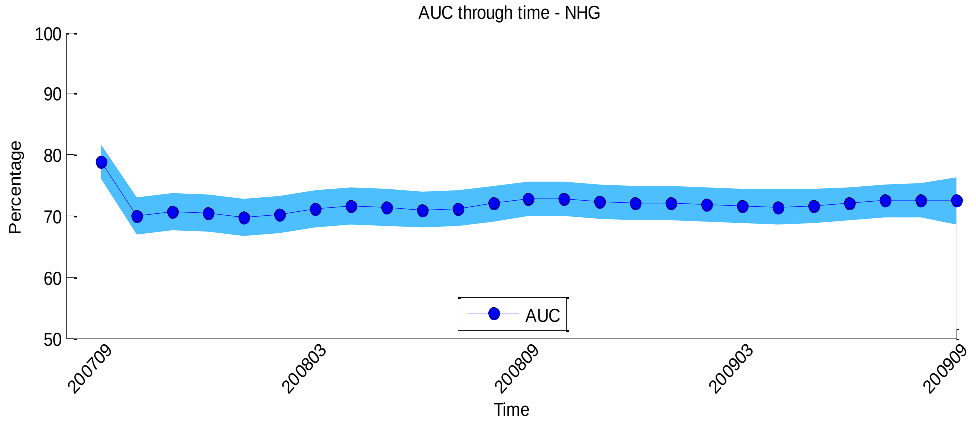


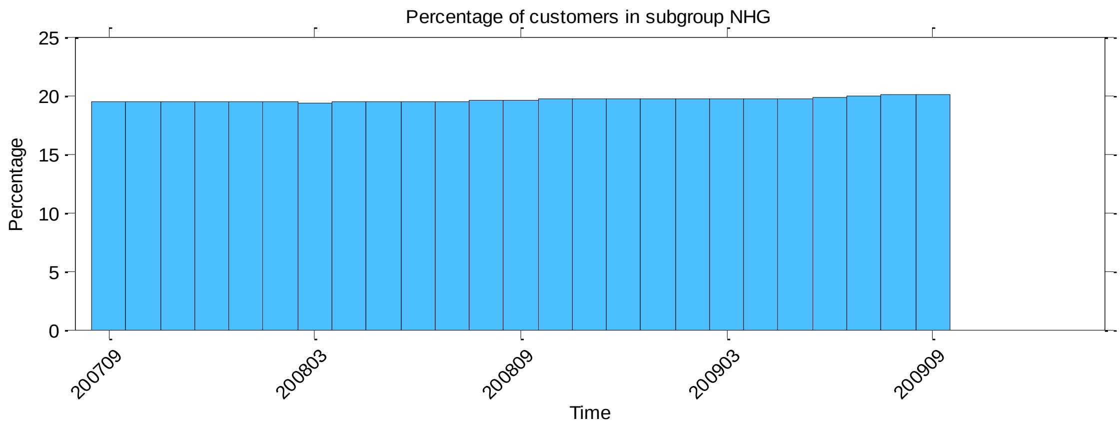


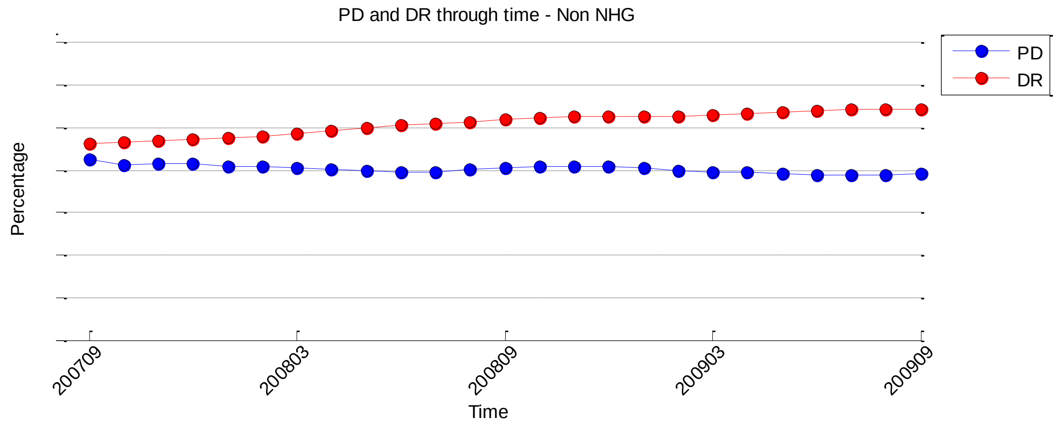


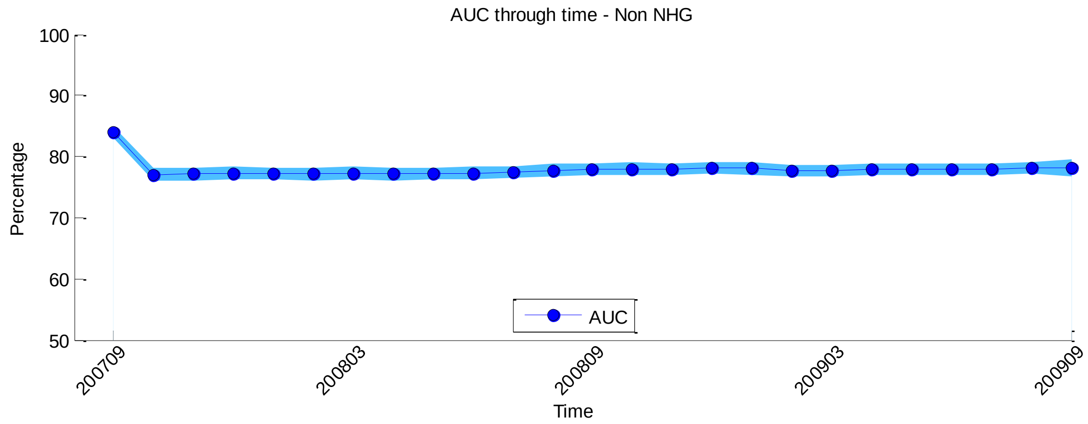
NHG/Non NHG

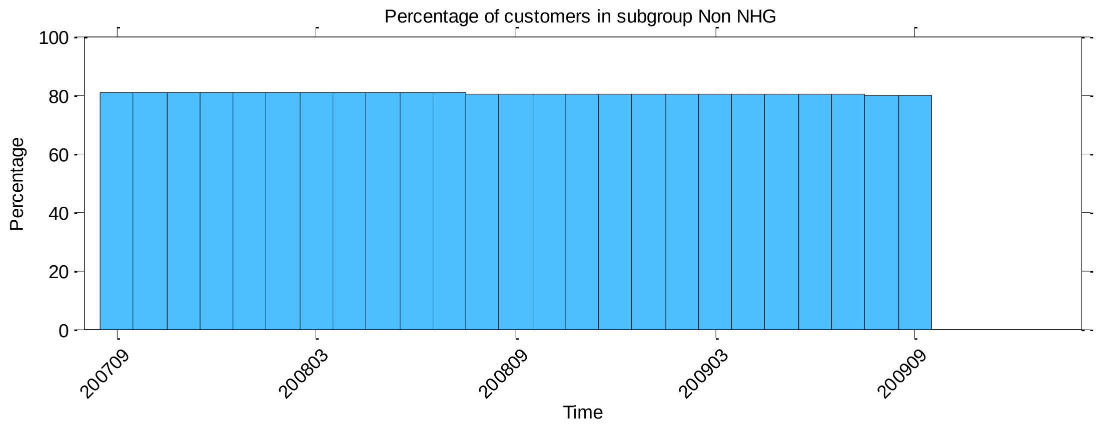

PD and DR through time - NHG
PD
DR
Percentage
200709
200803
200809
200903
200909
Time


AUC through time - NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

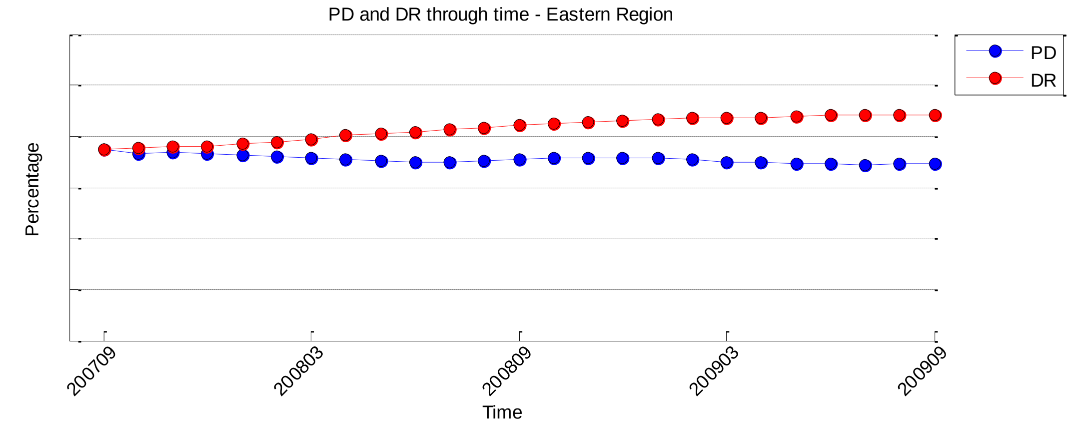

Percentage of customers in subgroup NHG
25
20
15
10
5
0
Percentage
200709
200803
200809
200903
200909
Time

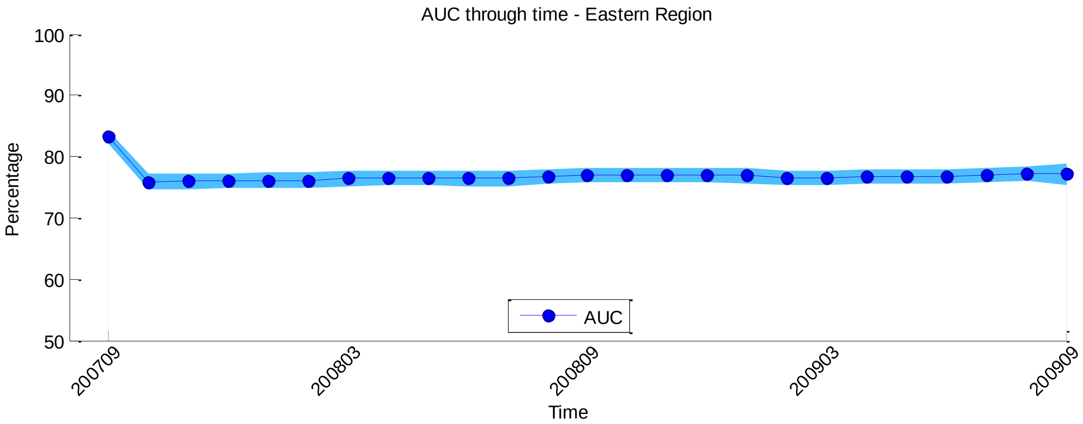

PD and DR through time - Non NHG
Percentage
PD
DR
200709
200803
200809
200903
200909
Time

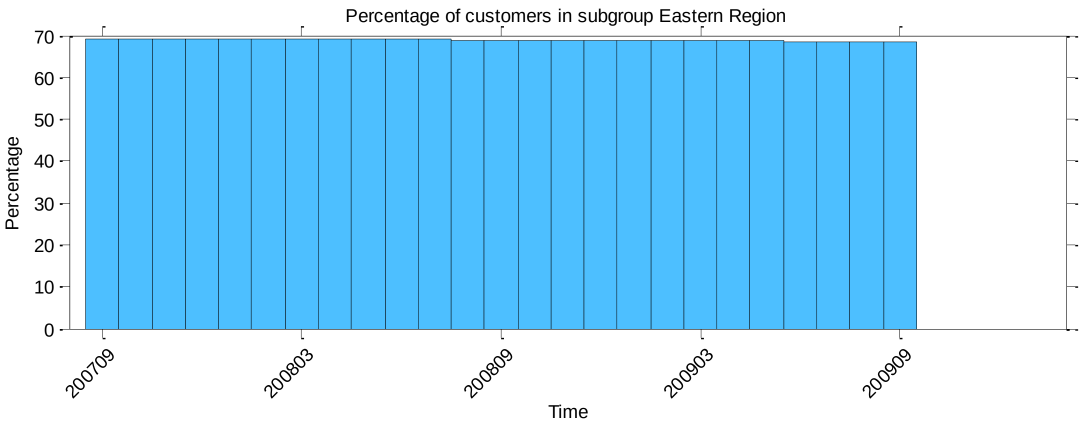

AUC through time - Non NHG
100
90
80
70
60
50
Percentage
AUC
200709
200803
200809
200903
200909
Time

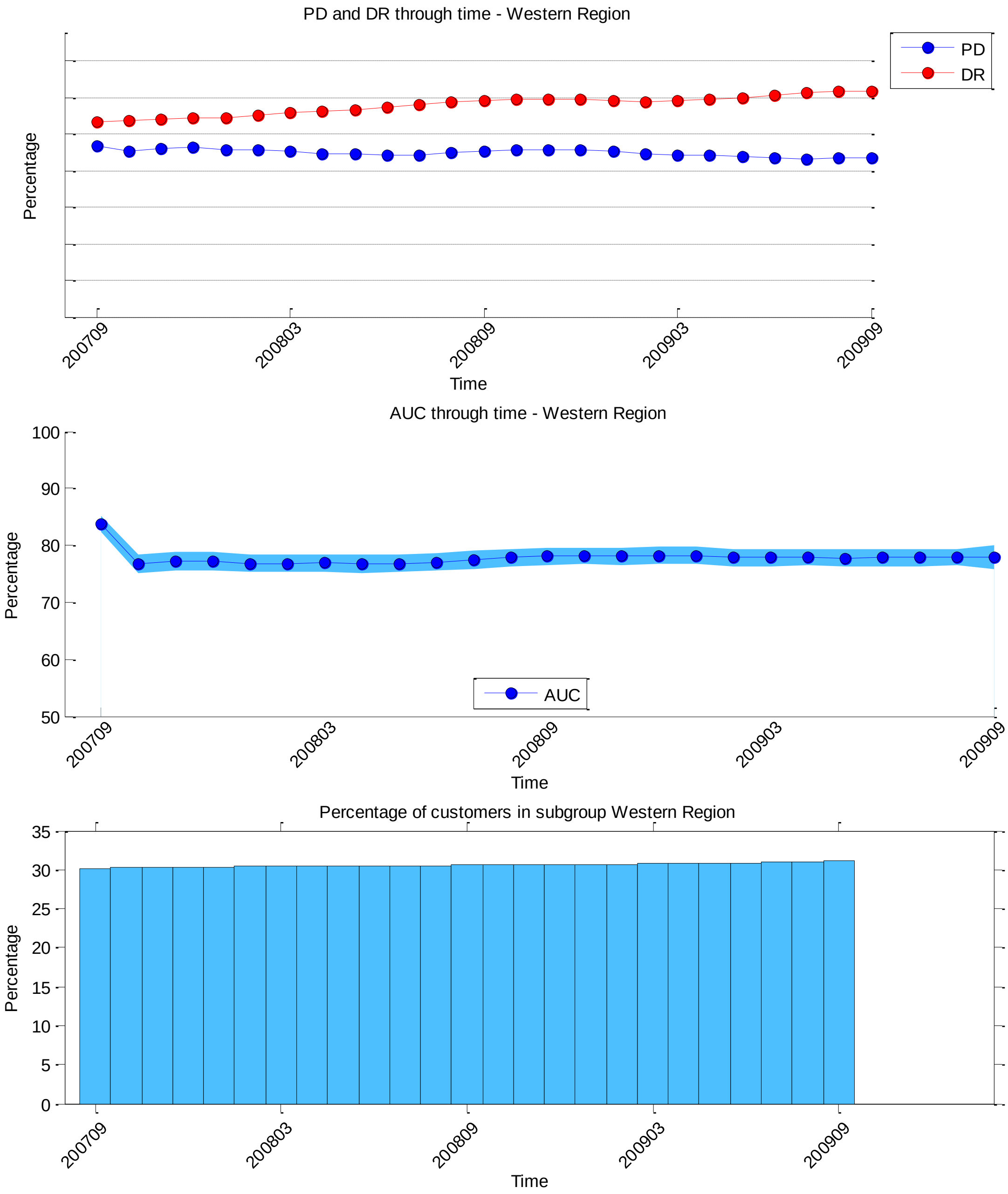

Percentage of customers in subgroup Non NHG
100
80
60
40
20
0
Percentage
200709
200803
200809
200903
200909
Time

Eastern Region/Western Region



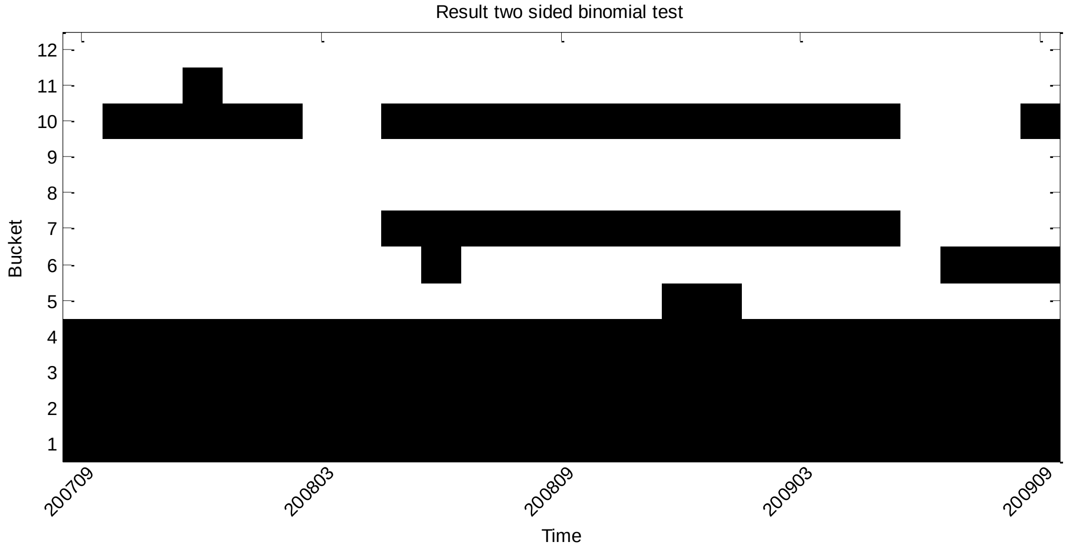




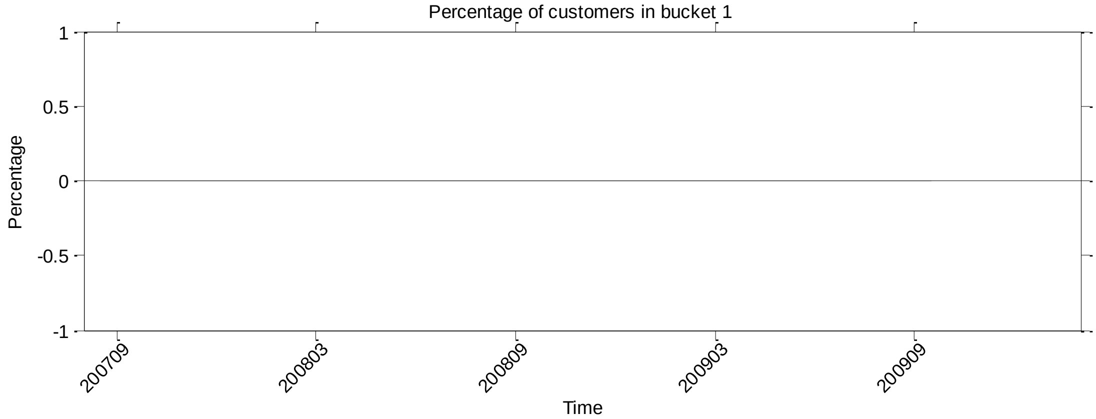

PD and DR through time - Western Region
PD
DR
Percentage
200709
200803
200809
200903
200909
Time
AUC through time - Western Region
100
90
80
70
60
50
Percentage
AUC
Time
Percentage of customers in subgroup Western Region
35
30
25
20
15
10
5
0
Time

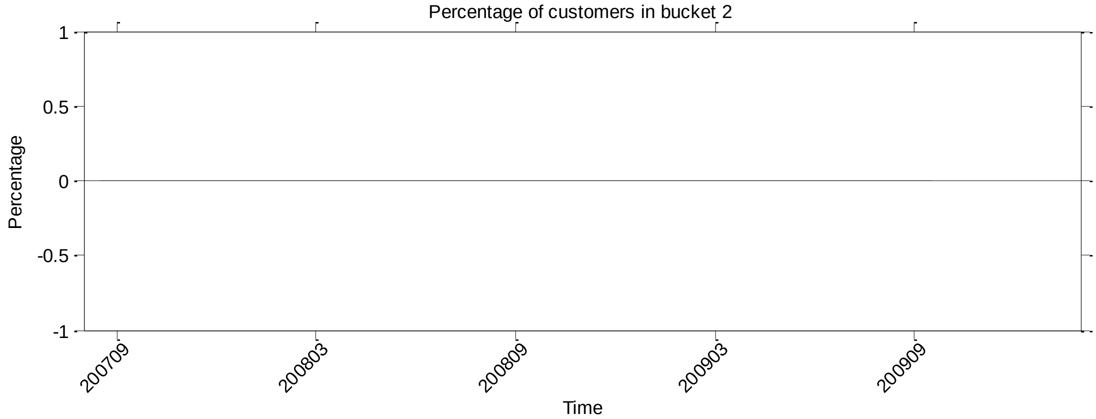
Binomial test Results on bucket level

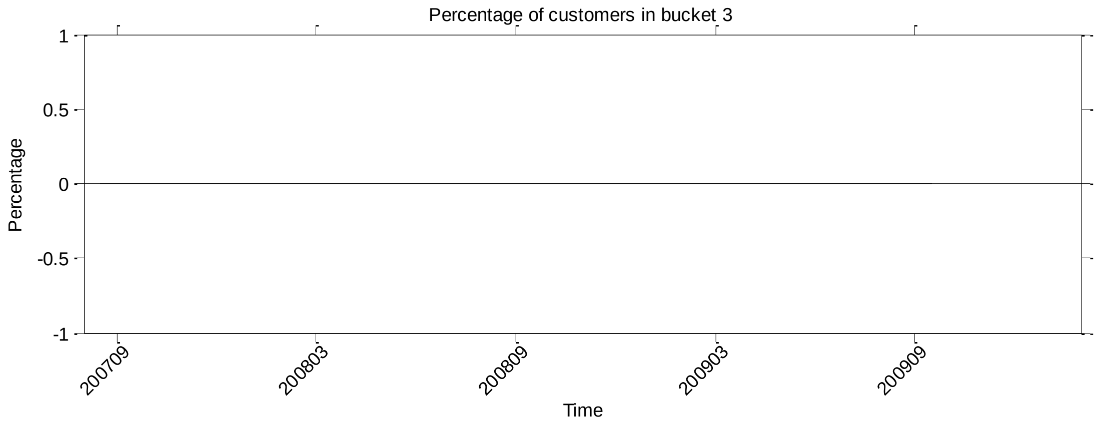
Important to notice is that buckets 1 till 5 and 9 contains together less than 2 percent of the customers. Because of the low number of observations in those buckets, the test results for those buckets cannot be seen as statistical relevant.

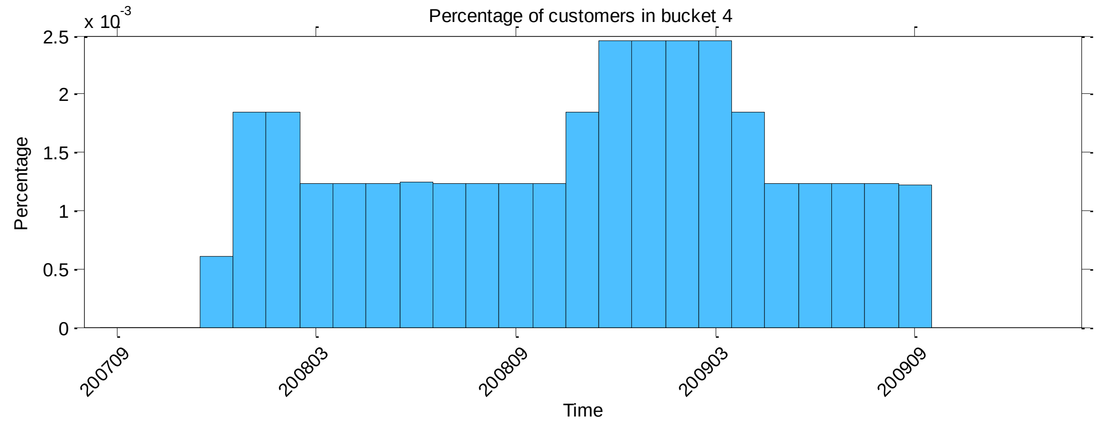


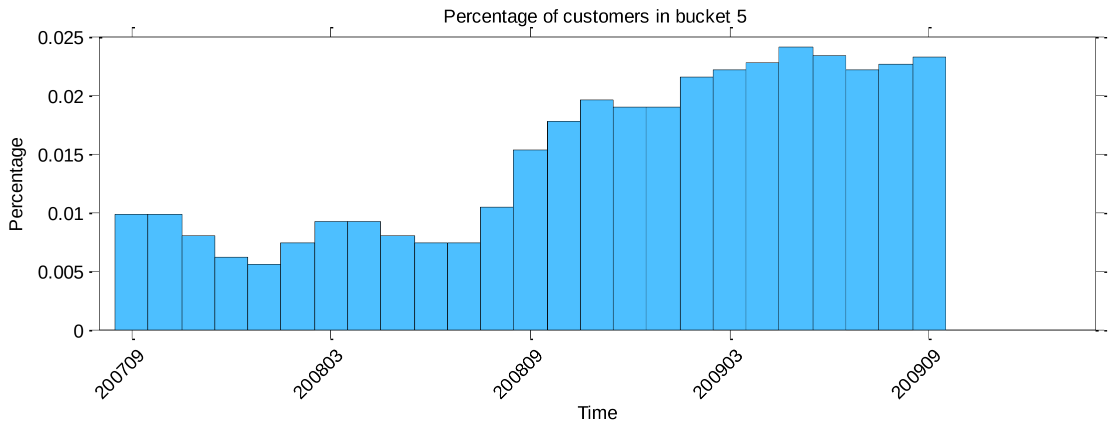
Percentage of customers per bucket

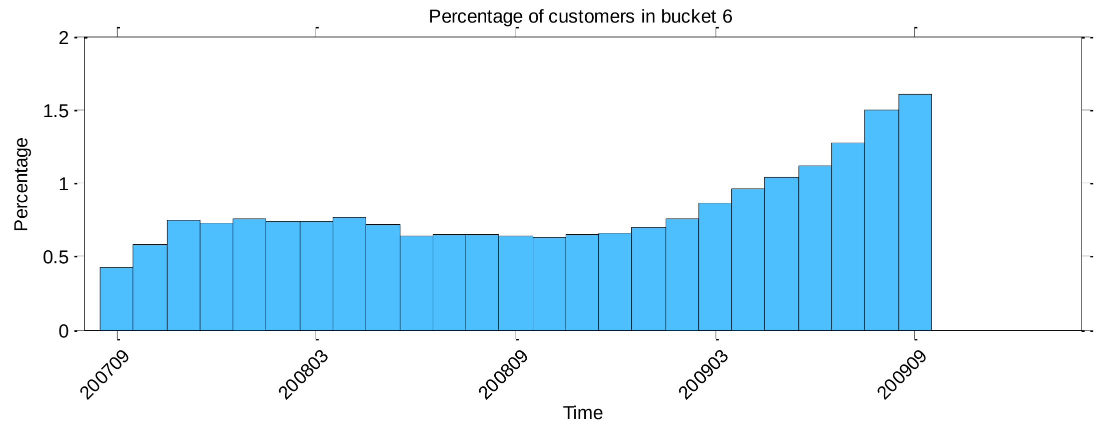


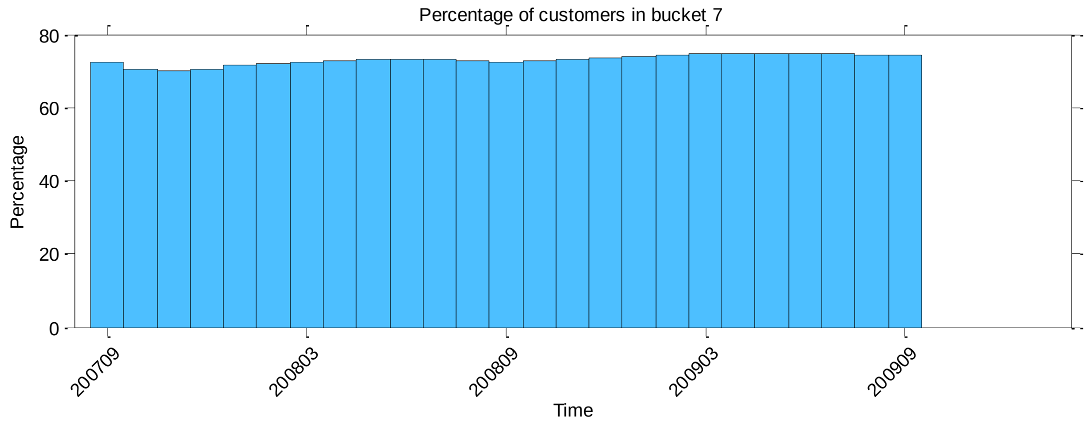


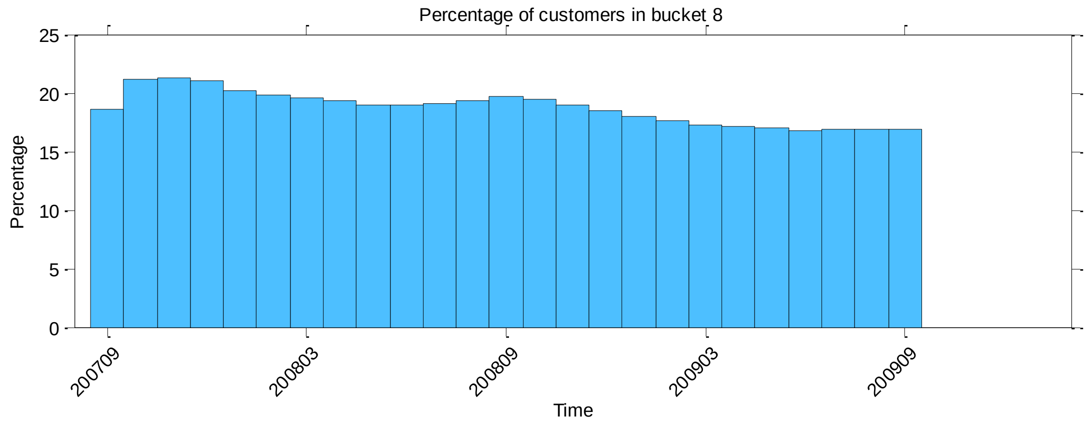

Percentage of customers in bucket 3
Percentage
1
0.5
0
-0.5
-1
200709
200803
200809
200903
200909
Time

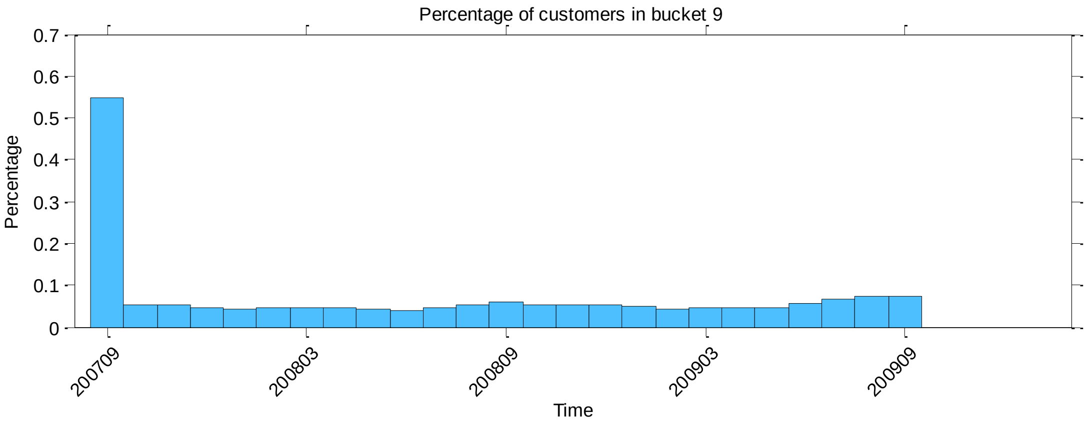

x 10^-3
Percentage of customers in bucket 4
Percentage
2.5
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time

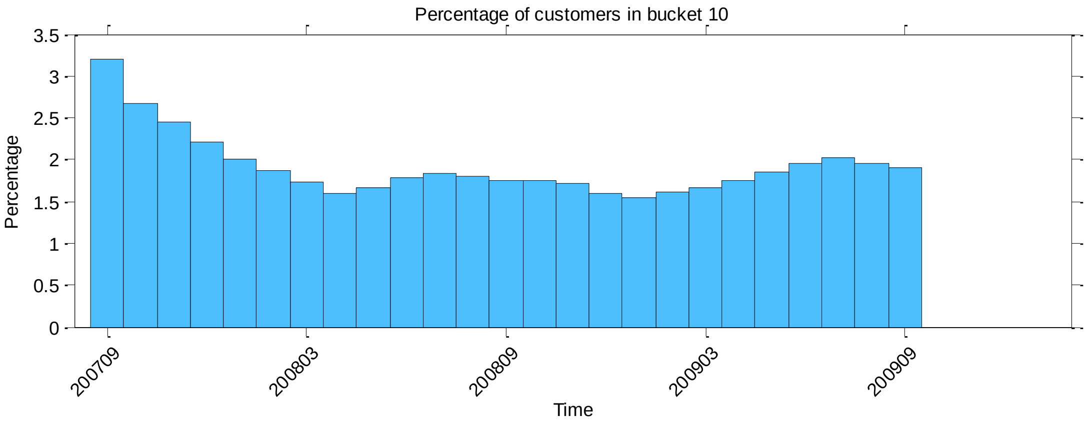

Percentage of customers in bucket 5
Percentage
0.025
0.02
0.015
0.01
0.005
0
200709
200803
200809
200903
200909
Time

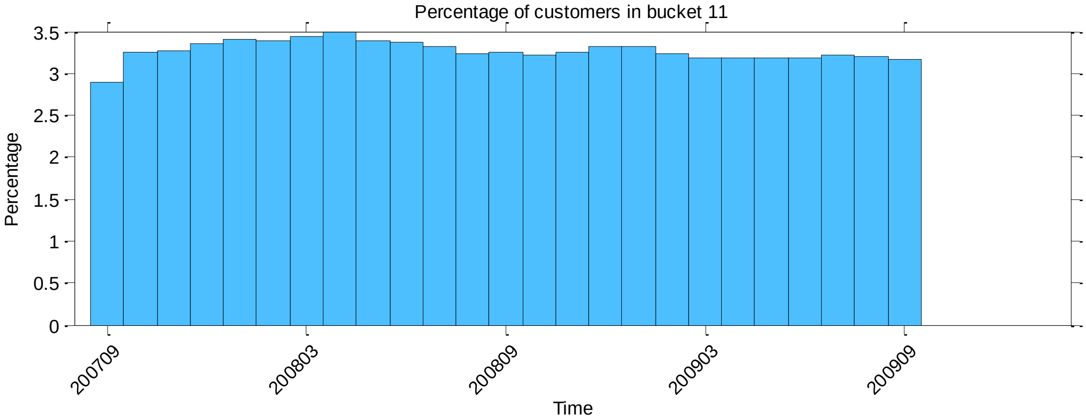

Percentage of customers in bucket 6
Percentage
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time

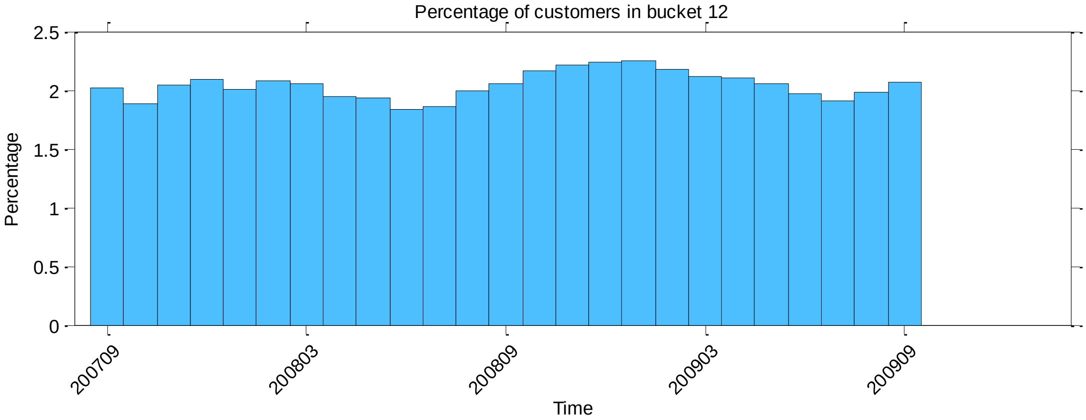

Percentage of customers in bucket 7
Percentage
80
60
40
20
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 8
Percentage
25
20
15
10
5
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 9
Percentage
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 10
Percentage
3.5
3
2.5
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 11
Percentage
3.5
3
2.5
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time


Percentage of customers in bucket 12
Percentage
2.5
2
1.5
1
0.5
0
200709
200803
200809
200903
200909
Time