Prepayment Risk in Residential Mortgage Loans

Degree Dissertation
for the
Master Examination in Economics and Finance
at the
Faculty of Economic and Social Sciences
of the
Eberhard Karls Universität
Tübingen

Examiner:
Professor Dr. Koziol

Submited by:
Konstantin Smirnov
Born in Nowokusnezk, Russian Federation

Date of submission: 23/02/2018

Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criterion
BGB	*Buergliches Gesetzbuch* - Civil Code of Germany
CPR	Conditional Prepayment Rate
$\widehat{CPR}$	Empirical Conditional Prepayment Rate
DTI	Debt-to-Income
EBA	European Banking Authorities
EU	European Union
FICO	Credit score created by the Fair Isaac Corporation
ICAAP	Internal Capital Adequacy Assessment Process
IRRBB	Interest Rate Risk in the Banking Book
LTV	Loan-to-Value
MSA	Metropolitan Statistical Area
SMM	Single Mortality Rate
$\widehat{SMM}$	Empirical Single Mortality Rate

List of Symbols

AIC	Akaike Information Criterion
CF_t	Realized Total Cash flow in month t
CPR_t	Conditional Prepayment Rate (cash flow) in month t
$\widehat{CPR}(t)$	Empirical Single Monthly Mortality Rate
D_E	Effective Duration
$h(t)$	Hazard Rate
$\hat{h(t)}$	Empirical Hazard Rate
H_0	H_0 hypothesis
i	Subjects or Loans
I_t	Monthly Interest Payment in month t
k	Region or MSA
$L(\beta)$	Likelihood Function of the parameter β
$LTV_i(t)$	Dynamic LTV for loan i
MP_t	Projected Monthly Mortgage Payment in month t
$MR_k(t)$	Averaged Mortgage Rate of region k
n	Original Term of the Mortgage in months
NR_i	Contractually agreed fixed note rate of the individual loan i
p	Number of Regression Parameters
$p(t)$	Probability Density Function
$P(t)$	Cumulative Distribution Function
PR_t	Prepayment Amount in month t
$PROP$	Value of the underlying property
PV	Present Value
r_{ik}	Schoenfeld Residual
$R(t_i)$	Set of Subjects at Risk prior time t_i
g_t^k	associated, extracted Case-Shiller growth rate at time t of MSA k
$S(t)$	Survival Function
$\hat{S(t)}$	Empirical Survival Function
SMM_t	Single Monthly Mortality Rate (cash flow) in month t
$\widehat{SMM}(t)$	Empirical Single Monthly Mortality Rate
SP_t	Scheduled Principle Payment in month t
t	Time or Month
$TAX.NY$	Estimated Risk Factor of the Dummy Variable "State New York"
$UNEMP$	Estimated Risk Factor of the Unemployment Rate
$YSPREAD_i(t)$	Yield Spread of loan i
$Z(t)$	Covariate
Δi	Shift of the Yield Curve, measured in Basis Points i
β	Parameter
$\hat{\beta}$	Estimated Parameter

1 Introduction

Nowadays the banking industry is facing significant challenges. Political insecurities and a sustained low-interest-rate environment pose difficulties to the financial system resulting in a decrease in overall profitability. In particular, this affects the bank's major core business: money lending.

Another critical challenge results from digitalization. Like any other business, the banking industry is primarily affected by strong technological development changing the industry's view of it self. Thereby, digitalization can be viewed as a threat, but also as a chance at the same time. Opportunities lay mostly in applying new methods using modern technological capacities. Thanks to improving computational power systems and effective large-scale data management, new possibilities emerge, extending the bank's customer experience.

Last, but not least, another key challenge arises from further regulatory requirements. Since the financial crisis in 2008, regulatory pressure still weighs heavily on the banking industry. Tightening capital requirements increase further costs, which are mainly derived from exposure to the market, liquidity and credit risk. Efficient risk management is therefore critical.

The focus of this thesis lays in the analysis of prepayment risk, which can be understood as part of both key challenges. In general, prepayment risk is derived from contractual or implicit optionalities, which give the debtor the right to prepay its unscheduled outstanding balance. First, it is of vital importance from a risk management perspective. Prepayment risk might strongly influence the expected cash flow of a debt agreement and therefore impact other risk sources such as market, liquidity and credit risk. In particular, these circumstances are taken into account in the just recently published European banking requirements addressing prepayment risk more concretely. Besides that, institutes are forced to handle emerging cost pressure more efficiently. Notably, the institutes' core business of lending money is affected by it. With the help of data analysis, expected prepayments can be better analyzed, optimizing the process of money lending.

All of these aspects are more important in a mortgage-market framework. Risk arising through private customers is mostly hard to capture and makes a behavioral statistical approach necessary. Due to missing market possibilities, this risk is not only hard to catch, but also hard to hedge.

Therefore, the objective of this thesis is to provide the reader with a theoretical overview of the topic of prepayment analysis in residential mortgages. In addition, these theoretical findings will then be explicitly applied in an empirical case study using mortgage

loan data.

The theoretical overview will first give the general definition of the term prepayment before new regulatory requirements emerging from the Basel III framework will be explained. Then basic methodological concepts of prepayments in mortgage markets and specific exposure on the institutes' risk will be discussed in more detail. Furthermore, an overview of factors driving prepayment risk will be provided to the reader.

Secondly, the empirical case study will be conducted on recent customer data of US mortgage loans provided by the Fannie Mae. As proposed by several authors such as Deng et al. (2000), Stepanova and Thomas (2002), Schultz (2015) and Banerjee et al. (2016) empirical prepayment rates should be measured using empirical survival methods. Also, it will be analyzed whether the discussed factors show statistical inference. One further goal is to evaluate whether empirical survival techniques can be applied in a German market environment from methodological and regulatory perspectives.

2 Theoretical and Regulatory Foundation

This section will firstly introduce the concept of prepayments from a financial economic and a regulatory perspective. The goal of this chapter is to provide the reader with a clear theoretical foundation on the concept of prepayments before the empirical analysis is conducted.

Therefore, as a starting point, the reader will be introduced to the general and juridical definition of the term *prepayment*. In order to highlight its contemporary importance of the topic of this thesis, the specific new regulatory requirements which arose from the actual Basel III framework will be disclosed. To prepare the reader for the empirical part, theoretical, methodological concepts will first be discussed. More specifically, the basic idea behind measuring prepayment rates and its emanating risk will be provided. Lastly, an overview of the different traditional economic but also socio-economic behavioral pattern influencing risk factors will be set to gain a deeper understanding.

2.1 Definition

A contractual debt obligation usually contains scheduled repayment cash flows in which the outstanding debt balance has to be paid back by the debtor to the borrower during the period of the loan or at some specified period. All repayments cash flows which appear unscheduled are defined as *prepayments*. Prepayments can occur either involuntary or voluntary. Involuntary prepayments occur whenever the debtor defaults,

leading to an involuntary full prepayment of the outstanding balance, usually financed by liquidation of the underlying asset of the debt contract. On the other hand, voluntary prepayments can occur due to contractual agreements. The debtor might have the right to exceed scheduled repayment cash flows to reduce future interest-rate costs or the debtor might even have the right to prepay the full outstanding debt obligation, by refinancing at a better rate. Both cases are referred to as voluntary prepayment options.[1]

Economic analysis of prepayments is strongly affected by the contractual agreement and therefore highly depends on the market it serves. Corporate debt agreements are structured in a different way than retail or mortgage contracts, and hence prepayment analysis might deviate sharply, as it will be briefly discussed in the later. Moreover, national legislation on prepayment exercisability plays a crucial role in prepayment analysis. Legal foundations do not only differ in the debt instrument itself but also differ profoundly among countries, especially when it comes to private loans.

For instance, for fixed-interest US mortgages, prepayment penalties do not exist. Therefore, the borrower can freely prepay his loan whenever possible, leading to a substantial prepayment activity and hence to a more complex risk environment from a bank's perspective. As a result, prepayment risk is of significant importance since it strongly affects valuation of the credit itself.

In Germany, however, prepayment exercisability might be firmly restricted by the statutory right of termination §489 BGB. For covered loans with a fixed-interest period less then ten years (e.g., usual mortgage loans), a full prepayment (cancellation) results in a prepayment penalty to cover resulting interest-rate losses, from a borrower's perspective. During that time these contracts usually contain only agreements of prepaying a specific percentage amount of the outstanding debt (generally up to 5% - 10% of the remaining balance). Therefore, in this case, exercisability is strongly restricted. Note that prepayment penalties are not allowed on floating loans. Also, thanks to the directive 2008/48 of the European Union, a prepayment penalty charged on uncovered consumption loans with volumes of less than 75.000 euro, was capped up to a maximum amount of 1% of the outstanding loan balance, giving the consumer more flexibility in prepaying. In this case, prepayments might affect the value more significantly. [2]

[1]The general term "prepayments" will be used hereafter to denote voluntary prepayments (as this is the focus of the master thesis).

[2]Note that these laws might not reflect legislation on corporate debt instruments which will not be explicitly discussed since this thesis focuses only on the mortgage market.

2.2 New Regulatory Requirements on the IRRBB

The newest regulatory requirements on prepayments options are addressed within the European Basel III (Pillar 2) framework.[3] Particularly, this topic is aimed within the draft guidelines of the European Banking Authorities on the *interest rate risk arising from non-trading book activities*, or in short *IRRBB* (EBA, 2017).[4] It results from the former standards of the Basel Committee on Banking Supervision (BCBS, 2016) on IRRBB which was already partly implemented at the EU level. The new draft guidelines on the IRRBB can hence be understood as an extended implementation of the existing legislation. The regulation will go into effect on the 31.12.2018.

The guidelines aim to substantiate identification, measurement, monitoring, and control of the impact of interest-rate movements on the bank's earnings and its effect on the market value of the bank's financial instruments in a short and long-term. Interest-rate risk exposure should be assessed in the risk management process and more importantly, be reported within the *Internal Capital Adequacy Assessment Process* (ICAAP).

Therefore, exposure on IRRBB should be considered within governance policies and in the institute's IT architecture. To identify all IRRBB related components, it requires banks to provide a full list of IRRBB-related instruments. As a specific subcomponent in that context, additionally, an inventory of all financial instruments exposed to *option risk* should be provided to the supervisory authority. In the context of prepayments, the EBA explicitly distinguished between prepayment options in mortgages contracts and further deposits, referred to as *behavioral options*, and prepayment options within wholesale instruments, referred to as *automatic interest-rate options*.

Furthermore, risk-exposure to IRRBB should be measured within the ICAAP by simulating different interest-rate shock scenarios. Depending on the size and complexity of the institute, simulations have to take different interest-rate scenarios into account in which behavioral and modeling assumptions on prepayments have to be considered.

Moreover, to take the exposure of behavioral options on the IRRBB into account, the institutes should evaluate the potential impact of prepayment speeds within different interest-rate scenarios. The various dimensions impacting prepayment behavior should be considered, whereby it is of particular importance that prepayment exercisability itself might be a function of the interest rate. All assumptions should be justifiable and accurate, explicitly concerning the historical data, and should be fully documented.

The empirical method used in this thesis could be understood as one possible approach

[3] It should be noted that further requirements on option risk are also addressed within the Financial Reporting Standards (IFRS), specifically in IFRS 9 and 13.

[4] These guidelines are a draft version and a subject to further change. However, as the past has shown, finalized versions contain only marginal differences.

to fulfill these guidelines. As it gets evident in the later, it could be interpreted as one way to model and justify behavioral assumptions on prepayment behavior for mortgages and similar private loans in a non-restricted prepayment environment.

2.3 Economic Foundation

2.3.1 Prepayment Rates

To understand the economic impact of prepayments in a behavioral option framework, fundamental concepts of annuity computation have to be first explained. An amortization table is the starting point of a prepayment analysis. It draws the cash flow structure of a pool of loans. First, the *projected monthly mortgage payment* or annuity has to be computed. It consists of the scheduled repayments and interest received from the debtor (Fabozzi, 2016):[5]

$$MP_t = MP_{t-1}\frac{i(1+i)^{n-t+1}}{(1+i)^{n-t+1}-1} \tag{1}$$

where i is the simple monthly interest-rate, n the original term of the mortgage in months, MP_t is the projected monthly mortgage payment for month t, and MP_{t-1} is the projected mortgage balance at the end of month t, *given prepayments have occurred in the past.* At this point, it is crucial to understand that annuity payments change monthly since prepayment change the annuity structure of the aggregated pool. Equation (1) has hence to be computed monthly concerning the new outstanding balance.[6]

The portion of monthly interest I_t of payment (1) can then simply be computed with:

$$I_t = iMB_{t-1} \tag{2}$$

The projected monthly scheduled principle SP_t is then simply the difference between the projected annuity payment (1) and (2):

$$SP_t = MP_t - I_t \tag{3}$$

The foundation of prepayment analysis is the *single monthly mortatlity rate* (SMM). In a pool of loans, it measures the monthly prepayment cashflows relative to the outstanding scheduled balance in the end of the payment period t. With respect to (1), (2) and (3), the SMM_t is then defined as:

$$SMM_t = \frac{CF_t - (I_t + SP_t)}{MB_{t-1} - SP_t} = \frac{PR_t}{MB_{t-1} - SP_t} \tag{4}$$

[5]Most concepts on prepayment in a behavioral framework are founded on the concept of US *mortgage-backed securities.*

[6]Indeed, this is a common mistake since it is often assumed that the annuity remains constant over time. However, this only holds under certain assumptions.

where CF_t is the realized total cashflow received in month t, and PR_t is the exceeding cash-flow amount or the actual prepayment amount received in month t.

An interpretation of the SMM could be: in month t, $\alpha\%$ of the outstanding mortgage balance available to prepay in month t, prepaid (Fabozzi, 2016). It is important to understand the SMM as a conditional concept. It measures the rate of prepayment in a cohort of loans, conditional of the outstanding amount of loans in that given payment period t. Or, expressed differently, it measures the *speed of prepayment* (Fabozzi, Schultz, et al., 2016).

The SMM in (4) is also converted to an annual level, defined as the *conditional prepayment rate* (CPR):

$$CPR_t = 1 - (1 - SMM_t)^{12} \tag{5}$$

For instance, a CPR of $\alpha\%$ implies that the approximately $\alpha\%$ of the mortgage outstanding loan balance will, besides the scheduled payments, additionally prepay at the end of the year.[7]

2.3.2 Measuring Prepayment Risk

The CPR impacts the net present value of the underlying debt obligation since interest-rate gains shrink due to prematurely repayments. Expected cash flows are therefore directly influenced, resulting in an additional risk factor which has to be taken into account. The key concept in measuring its impact on the asset's economic value is the *effective duration*. It measures the interest-rate sensitivity of the asset's net present value, in case of a parallel shift in the underlying yield curve (discount curve). Compared to traditional duration concepts (e.g., Macaulay duration), the effective duration is capable of taking expected cash flows into account which is crucial in prepayment analysis (Fabozzi, Buetow, et al., 2005).

The effective duration D_E is defined as (Choudhry, 2010):

$$D_E = \frac{PV_- - PV_+}{2PV_0\Delta i} \tag{6}$$

where

PV_0 = the estimated present value of the loan under the current yield curve

PV_- = the estimated present value of the loan under a parallel upward yield shift of Δi

PV_+ = the estimated present value of the loan under a parallel downward yield shift of Δi

Δi = the shift on the yield curve, measured in basis points i

[7] An exemplary amortization table is provided in the Appendix.

As an example, the effect of three different prepayment scenarios (no CPR, 2% and 3%) on (6) in a 1,000,000$ pool of 30-year fixed-income mortgage loans is illustrated in figure 1.[8] Under the current upward-sloping yield curve regime, the present value of the optionless pool would change by ± 13.92% in a parallel downward or upward scenario. Assuming an annual prepayment speed of 2%, the present value would react only by ± 11.75%, and for a 5% CPR, the present value would change only by ± 9.28%. Or stated differently, in a parallel shift scenario, the present value of a pool of loans with an annual prepayment speed of 2% or 5% , would react 16% or 33% less sensitive, compared to the no CPR pool. The reason for this are the different cash flow term structures. In the CPR-case, payments agglomerate due to earlier prepayments in the short-periods of the pool's durability. Under the upward-sloping yield curve regime, long-term future cash flows are therefore weighted less, resulting in a general less-sensitive asset.

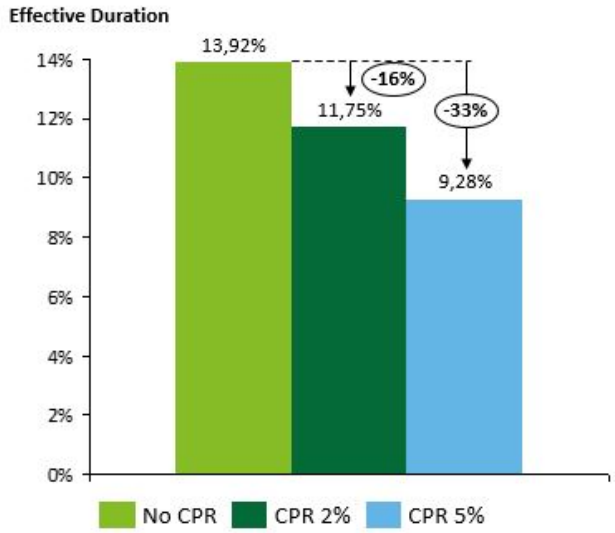


Figure 1: Effective Duration

However, a static CPR ignores that exercisability depends on the yield curve itself. Indeed, assuming that lenders behave rationally, refinancing incentives would arise whenever interest rates shrink to a marginally lower refinancing rate. Therefore, if prepayment speed is an endogenous function of the interest rate itself, the net present value would react less sensitive to a yield curve drop, as illustrated in Figure 2. In that example, a 1% yield change would result in positive 15% present value change assuming that the borrowers would adjust their prepayments to the new interest-rate environment dynamically. In this scenario, the present value of the pool would only positively change by 5%. In the financial literature this phenomena is referred to as

[8]The illustration was computed with an Excel-based interest-rate risk calculator developed by *Deloitte*. All computations are attached within a separated Excel file *"Duration"*.

negative convexity (Fabozzi, Buetow, et al., 2005). [9]

On the other hand, a rising yield-regime would reduce prepayment speed to a minimum. From a borrower's perspective, scheduled necessary payments would be the most rational choice since the relatively low agreed mortgage-rate could be 'locked in'. In order to prevent opportunity losses, surpluses should rather be reinvested in the relatively high riskless rate. As illustrated in Figure 2, a constant CPR might therefore overestimate interest-rate loses. Exercising prepayment option should therefore at least be understand as an endogenous function of the yield curve itself. From a risk perspective, ignoring endogeneity in a growing yield-regime would result in an underestimation of the interest-rate risk, while in an inverse yield regime, interest-rate risks might be overrated. From that perspective, an endogenous prepayment model is inevitable.

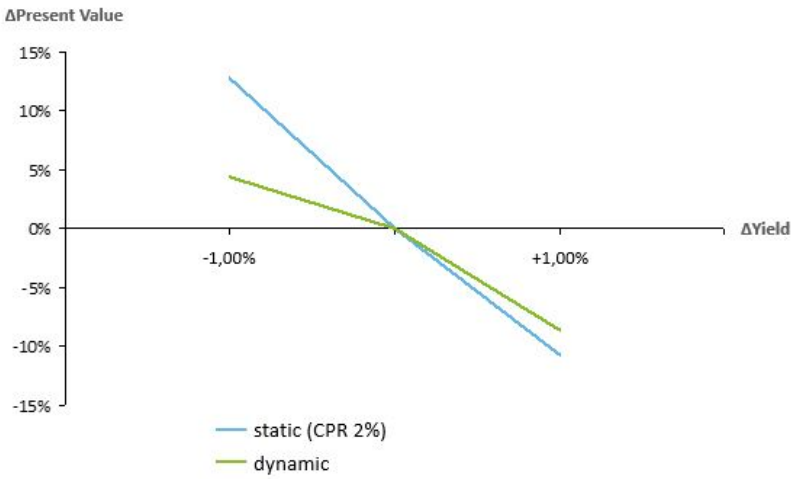


Figure 2: Negative Convexity

2.3.3 Outlook: An Option-Pricing Framework in Prepayment Analysis

The speed of prepayment should rather be understood as a function of the yield curve itself. Indeed, for different corporate debt-instruments, the refinancing incentive is usually assumed as the only driver of prepayments. Note that due to this strong assumption, prepayment analysis in this market segment profoundly differs to the mortgage market, which is the focus of attention of this thesis. However, for the sake of completeness, it is useful to introduce the reader shortly to the concept of corporate debt analysis.

[9] Note that in this illustration it was assumed that a 1% yield shock would result in a rise of CPR to 5% while a negative yield shock would lead to a decreasing CPR rate of 1%. The baseline CPR was assumed to be 2%.

In the corporate market, analysis and valuation of prepayment options are usually conducted in an option-pricing framework. Indeed, full prepayment can be understood as a call option, in which the debtor can terminate his debt contract or bond by refinancing for lower interest rates. Stated differently, the call option will be exercised if the spread between the agreed interest rate of the agreement and the observable refinancing rate itself is positive or *in the money*. From a financial perspective, the borrower is therefore long (and the lender short) in a call position in the underlying asset.

This issue is already widely discussed in the option-pricing literature in which debt or bond contracts with prepayment features are referred to as *callable bonds*. These options are usually valued in a risk-neutral framework using different numerical methods to solve for the underlying differential equation (e.g. see, Brennan and Schwartz (1977) , Hull and White (1990a) or Hull and White (1990b)). However, option pricing models rely heavily on its assumptions. Different models try to loosen or extend different assumptions, e.g., models extend the underlying interest-rate term structures and exercising possibilities.

When it comes to mortgage contracts, many assumptions are hard to justify. The essential assumption of risk-neutrality implies pure rationality of the agent's exercising behavior. In a classic option-pricing-framework prepaying is only related to its refinancing incentive, suggesting that even a marginal change in the observable refinancing rate would lead to an exercise of the option. This pure rationality assumption might hold for large corporations like banks with specialized market departments, but in a simple mortgage framework, this assumption is hard to justify. The exercise of the option of a 'simple' agent is not only dependent on its refinancing incentive. There are more factors which have to be considered regarding the exercise probability compared to a callable bond, resulting in a different valuation technique. Indeed, as discussed in chapter 2.2, the EBA differentiates between automatic options on wholesale credits and behavioral options in mortgages.

Applying the rational option-pricing framework to value mortgage prepayment is therefore not appropriate since it would neglect essential behavioral and economic patterns of the agents. An overview of these factors will now be given in the next chapter.

2.4 Prepayment Drivers

As it gets clear, not only is an accurate measurement and analysis mandatory from an economic perspective, it is also important from a regulatory point of view. In that sense, for extended analysis, it is not only important which actual risk is linked to a prepayment right, but also what specifically is driving prepayment exercisability of the

underlying asset. Following the same logic as the EBA (2017), exercisability is not only driven by the yield curve, as it is the case for a callable bond. In fact, a mortgage contract can be understood as a callable bond with a complex option structure which attempts to minimize the market value of the option's liability but exercised inefficiently (Weiner, 2016). This section will give an overview of other risk factors which influence prepayment in mortgage contracts according to financial literacy. This section can be seen as the preliminary work of the empirical framework in the following chapter since most of the stated findings are directly considered in the analysis of this thesis.

It is essential to make clear, how modeling patterns have developed away from an aggregate level to a more granular approach (Banerjee et al., 2016). Due to a lack of loan-specific data (loan-level data) 'old' empirical models mostly analyzed aggregated cash flows, neglecting heterogeneity of the underlying loans. Therefore, prepayment functions were mainly understood as macroeconomic driven processes. Thanks to current-day technology, vast amounts of loan-level data are collected and provided for research purposes. Modern modeling approaches hence incorporate loan-specific effects to address heterogeneity, resulting in a more accurate approach. Furthermore, nowadays data is analyzed on the most granular level possible, e.g., mortgage rates based on geography or unemployment rates of the MSA (Metropolitan Statistical Area) are preferred to national data, resulting in more precise risk estimation.

2.4.1 Macro drivers

The incentives of refinancing are still regarded as the most important factor driving prepayments. In the economic literature, it is addressed in various ways, starting from a simple empirical model, in which the refinancing incentive is simply modeled as the spread between the agreed mortgage rate and the observable mortgage rate in that period, to highly complex mathematical incentive functions.

Schwartz and Torous (1989) model the refinancing incentive of a loan-pool as a time-varying cost function, in which the spread between the average mortgage rate of the pool and the long-term treasury rate drives the refinancing costs. Most importantly, they regarded the refinancing variable as a lagged function over several months, resulting in more accurate results. Stanton (1995) tries to model refinance incentive as the value of the mortgage holders liability, which is represented as an extended stochastic interest-rate differential equation. Another popular approach was presented by Deng et al. (2000). In their approach, the refinancing incentive is modeled as a call option in the spirit of the traditional derivative framework. Their model is similar to a callable bond framework, whereas it additionally incorporates further private mortgage-related variables like house-prices or outstanding balance. In general, their model can be described as a mix of classic derivative-pricing and empirical modeling. A recent overview

of different approaches to capture refinancing incentive is presented in Schultz (2015).

An essential phenomenon which is related to time aspects of refinancing is the so-called *burnout* effect. It could be shown that in a regime of in-the-money mortgage rates, prepayment in a pool of loans slows down with increasing time, even though strong refinancing incentives exist. This phenomenon is a result of an adverse selection process. If refinance incentive exist, 'healthy' loans with the capability of refinancing will prepay as soon as possible. Therefore with increasing time, only loans with poor prepayment performance are left in the pool leading to a slow-down in prepayment (Fabozzi, Schultz, et al., 2016).

The second most crucial macroeconomic driver is home price appreciation, also referred to as the *turnover rate* (Joshi et al., 2016). It is of specific priority whenever loans are out-of-money from a mortgage-rate perspective. Whenever home prices appreciate in value, selling the property and prepaying outstanding debt might be a reasonable option, even if refinancing incentive do not exist. Turnover is also time-dependent since selling the property, in general, is profitable in the midterm. In modern prepayment analysis, turnover rates usually incorporate the granular level, specifically using the loan to value (LTV) ratio. The discussion on turnover rates will be therefore extended in the LTV part of that chapter.

Furthermore, traditional macroeconomic variables like unemployment influence exercising probability (Deng et al., 2000). In theory, high unemployment has a negative impact on prepayments because the capability of refinancing is reduced on average over the whole pool. Loans which suffered a job loss have no access to refinancing, and it is likely to assume that individual excess equity is not used to reduce debt. On the other hand, low unemployment rates indicate a healthy economy which might come with higher wages and hence higher individual equity accumulation which might increase prepayment speed. Another 'macro' factor is the divorce rate which positively impacts prepayment since it might lead to a forced housing-turnover.

Also, geographic location of the borrower itself might also play a major role (Banerjee et al. (2016)). In certain states, different socioeconomic patterns drive prepayment rates. For instance, the population of California traditionally is known for its high mobility, thereby leading to increased turnover rates. Additionally, state policies might play a crucial role, e.g., in the state of New York, a refinancing tax exists which negatively impacts prepayment speed naturally. In general, these effects can be incorporated with the help of a geographical variable.

2.4.2 Loan-level drivers

One important loan-level factor is a credit-driven prepayment (Weiner, 2016). In general, a high credit score borrower has a higher chance of prepaying. These loans have better access to refinancing opportunities and could accumulate excess equity more easily, while for low credit score borrower, refinancing alternatives are not available. Also, prepayment incentive might arise due to an improved, updated credit score, allowing the borrower to refinance at better rates.

The most influential variable which affects voluntary prepayment is the loan-to-value ratio (LTV) (Banerjee et al., 2016). It relates the loan amount to the value of the property measuring how much of the property value can be covered by the outstanding debt when the property would be liquidated. For example, an LTV of 50% indicates that only 50% of the property equity is necessary to repay the outstanding debt, while a rate above 100% would suggest that the value of the property could not cover the outstanding balance. As briefly mentioned, this measure is closely related to the turnover rates since it combines house-price appreciation and the individual outstanding balance. Therefore, it should not be viewed as static, time-fixed origination variable because property development and the borrower's debt level might strongly vary over time.

The effect on prepayment is similar to turnover rates. Favourable developing housing prices and (or) decreasing debt obligation lead to a decrease in LTV from which refinancing opportunities might arise (and vice versa). Specifically, it increases the probability of a prepayment transaction referred to as *cashout refinance*. Here, not only is the property value used to refinance the outstanding obligation at a lower rate but the borrower also asks for additional debt to free up low-rate liquidity funds. These funds can then be used to prepay other more expensive outstanding debt, e.g., car loans, or be used for further consumption purposes (Joshi et al., 2016).

Besides that, LTV might also indicate mortgage-insurance related prepayment events. In the US, it is obligatory to provide a mortgage insurrance for debts with ratios above 80%. Due to the "homeowners protection act 1998", borrowers have the right to cancel loan insurance by requesting a revaluation of the mortgage. Given that opportunity, borrowers might alternatively choose to refinance at lower rates by canceling. On the other hand, an increase in LTV impede prepayments due to a lack of refinancing alternatives (Banerjee et al., 2016).[10]

Another property-related driver is the occupancy status of the property. A borrower might occupy the mortgage or utilize the property as a purely financial investment.

[10]Indeed, growing LTV ratios are one of the most influential drivers of involuntary prepayments and credit events (Weiner, 2016).

It can be suggested that financial investment loans tend to prepay faster due to an increase in refinancing capabilities or simply due to a broader market understanding of the loan agent.

Also, personal loan-related information profoundly impacts prepayments behavior. For instance, it is reasonable to assume that the individual age of the agent might be linked to prepayment behavior. Older agents, on average, might have accumulated merely more equity over their lifetime, which increases the exercise probability and vice versa. A more general personal financial variable is the debt-to-income ratio (DTI). It relates the monthly required debt payment to the monthly income of the agent (Schultz, 2015). Higher income might indicate excess liquidity funds, which might be used to repay the loan. In theory, lower DTIs are linked to higher prepayments and vice versa. Note that highly personal data about the monthly income of the lender might not be easy to capture because a borrower usually does not have to update the lender with actually income statements over the time of the loan. Typically, the only events that have to be reported are negative 'trigger' events.

The loan-volume itself might also be considered as a further prepayment driver (Banerjee et al., 2016). A large loan-volume indicates higher prepayments speeds since interest-rates costs are relatively high compared to small loan amounts. Therefore prepaying the loan could decrease interest rate costs. Besides this, people, in general, avoid being in debt. Debt aversion results from psychological aspects leading to an increase in prepayment speed, even though it might be rational to be in debt. On the other hand, refinancing might not be a reasonable option for smaller loan amount due to the incurred relative fixed-costs in the transaction process.

2.4.3 Time-Effects and Overview

Moreover, time factors might affect refinancing decisions, e.g., seasoning patterns might influence prepayment behavior. Schwartz and Torous (1989) assume a seasoning pattern, in which prepayments occur more likely in summer months than in winter months. An explanation could be that people tend to move more often in the summer break due to personal time preferences. Bear in mind that, as already discussed, national jurisdiction restrict prepayment policies, making time as one of the most integral aspects of prepayment analysis. For instance, it is reasonable to assume that the discussed factors of this chapter are of subordinated importance for the German market due to the implicit ten-year prepayment restriction. Prepayment drivers might only impact exercisability as soon as the statutory period of notice §489 *BGB* applies.

Also, note that prepayment drivers were only discussed from an isolated perspective. Multiplicative effects between the drivers were ignored in this discussion. Of course,

these effects exist, e.g., low credit-score might counterbalance a high LTV ratio and vice versa. However, the purpose of the master thesis is to analyze the statistical inference and its marginal impact. Multiplicative effects might play a more decisive role when it comes to predictive model building. An overview of all discussed prepayment drivers is given in the following Table 1.

Table 1: Prepayment Drivers: Overview

Macro	**Loan-Level**	**Time**
Mortgage Rates	Credit-Score	Years
House-Price Appreciation	Loan-to-Value (LTV)	Seasonal Patterns
Unemployment	Loan Volume	
Geography	Debt-to-Income (DTI)	

3 Empiricism

The theoretical framework provided a starting framework for the empirical case study conducted in this chapter. This part of the thesis can hence be understood as an empirical extension on the discussed aspects.[11] As proposed by Deng et al. (2000), Stepanova and Thomas (2002), as well as Schultz (2015) and Banerjee et al. (2016) a survival analysis was applied on historical loan dataset provided by Fannie Mae. This approach might be also regarded as an application of the new EBA (2017) guidelines from Section 2.2. Additionally, this can be used to deepen customer understanding in the banking core business of money lending.

In the first chapter of the empirical part, the reader will be introduced to the methodological foundation of survival analysis. The basic empirical methodology will be additionally extended to the popular Cox model (Cox, 1972) to analyse the impact of selected risk drivers from chapter 2.4. The second subchapter will provide a detailed description of the data set in general and summarizes the analyzed variables which were considered in the Cox regression. Lastly, the empirical results will be presented to the reader, in which also the power of estimated prepayment rates will be compared to the cashflow counterparts in equation (4).

[11] Note that all following computations, estimations and illustrations were conducted using R-Studio v 1.1.423.

3.1 Methodology

3.1.1 Survival Analysis Fundamentals

Under survival analysis, a large variety of statistical methods are understood whose goal is to examine and model the time of a random event to occur, given a period of study. The terminology might be misleading in a prepayment analysis, but this comes from the fact that it was first applied in the field of biology and medicine in which usually the time to death was examined.[12]

One major beneficial property of survival analysis compared to traditional econometric approaches is the incorporation of *right censored data*, a peculiar feature of time-to-event data (Klein and Moeschberger, 2005). In the context of prepayment analysis, a cohort of loans is defined to be right-censored whenever a loan has "survived". The date of termination lays beyond the study period and therefore on the "right-side" of the time-axis. Right-censoring can also occur if a loan drops out of the data set without an event occuring, e.g., a loan just might got lost due to technical mistake.[13] Another aspect might be "left-truncation" or delayed entry. It describes subjects which entered the data set after the defined starting time (Schultz, 2015).

More formally, let T be the time until the event of prepayment. Assuming that T is a non-negative random variable from a homogeneous population, the distribution can be characterized by the survival function, the hazard function, and the probability density function. All three define another representation of the survival distribution in the group of study.

The cumulative distribution function of T is described as:

$$P(t) = Pr(T \leq t) \tag{7}$$

the probability density function is then defined as:

$$p(t) = \frac{dP(t)}{dt} \tag{8}$$

Assuming that T is continuous, the survival function $S(t)$ is then the complement of the distribution function (7) and describes the probability of a subject surviving beyond time t:

$$S(t) = Pr(T > t) = 1 - P(t) \tag{9}$$

[12] Survival analysis is sometimes also referred as "event-history analysis" in sociology or "failure-time analysis" in engineering (Fox, 2002).

[13] Other forms like left-censoring or interval censoring are not relevant in the context of prepayment analysis and will, therefore, not be discussed.

Note that any survival function shares the same property: it is a monotonic, non-increasing function equal to 1 at t=0 and zero with time approaching infinity.

Another representation of the survival time is the hazard rate. It describes the instantaneous risk at time t, conditional that the subject is alive just before that time. Assuming that T is continuous, the hazard rate can expressed as (Klein and Moeschberger, 2005):

$$h(t) = \lim_{\delta t \to \infty} \frac{Pr[(t \leq T < t + \delta t) | T \geq t]}{\delta t} \tag{10}$$

In the prepayment framework, T is a discrete (monthly), random variable. Suppose that T takes on discrete values t_j, j=1,2,.. , given a probability mass function (Klein and Moeschberger, 2005):

$$p(t_j) = Pr(T = t_j) \tag{11}$$

With respect to (11), the discretized survival function can then be described as:

$$S(t) = Pr(T > t) = \sum_{t_j > x} p(t_j) \tag{12}$$

The survival function can then also be expressed as a product of conditional survival probabilities:

$$S(t) = \prod_{t_j \leq t} \frac{S(t_j)}{S(t_{j-1})} \tag{13}$$

Note that in the discrete case, survival is as non-increasing step-function.

For $j = 1, 2, ..$, the discretized version of the hazard rate is defined as:

$$h(t_j) = Pr(T = t_j | T \geq t_j) = \frac{p(t_j)}{S(t_{j-1})} \tag{14}$$

where $S(t_0) = 1$. Since (11) can be described as: $p(t_j) = S(t_j) - S(t_{j-1})$, the hazard rate, with respect to (14) can also be described as:

$$h(t_j) = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} = 1 - \frac{S(t_j)}{S(t_{j-1})} \tag{15}$$

Note that with respect to (13) and (15), the hazard rate is related to the survival function in following fashion:

$$S(t) = \prod_{x_j \leq x} [1 - h(x_j)] \tag{16}$$

In a prepayment framework, the empirical hazard rate is described as the instantaneous risk that a loan prepays at time t, given the probability that the loan is still alive just before t, with respect to (15):

$$\hat{h(t)} = \frac{N(t_{j-1}) - N(t_j)}{N(t_{j-1})} = \frac{D(t_j)}{N(t_{j-1})} \tag{17}$$

where N is the number of loans at time t_j or t_{j-1}, and P is the number of loans prepaid in the time interval j.

With respect to (16) and (17), the empirical survival function for a cohort of loans can then be described as:

$$\hat{S(t)} = \prod_{x_j \leq x} [1 - \hat{h(t)}] \tag{18}$$

(18) is also referred to as the non-parametric *Kaplan-Meier* or the *product limit estimator*.

It is crucial to understand that (17) is closely related to the defined SMM (4) or CPR (5) of Chapter 2. Whereas these concepts measure prepayment on a cashflow level conditional on the outstanding balance of a pool of mortgages, the empirical hazard rate measures the count of loans failing conditional on the outstanding number of loans in the pool. Some researchers even claim that both concepts are indeed equal (e.g. (Deng et al., 2000) or (Schultz, 2015)). In the further analysis, it is therefore assumed that the Kaplan-Meier estimates of the empirical hazard function (17) are equal to the SMM in (4). The SMM is then defined as the empirical $\widehat{SMM}(t)$:

$$\hat{h(t)} = \widehat{SMM}(t) \tag{19}$$

and therefore with respect to (5) and (19), the empirical $\widehat{CPR}(t)$ can then be defined as:

$$\widehat{CPR}(t) = 1 - (1 - \widehat{SMM}(t))^{12} \tag{20}$$

Equations (19) and (20) are particularly in the context of the new EBA (2017) useful results. However, as we will see, this only holds when it is assumed that prepayments are driven mainly by full prepayment or loan contract cancellation. This claim will be analyzed in more detail in chapter 3.3.1.

3.1.2 Cox-Proportional Hazard Model

The Kaplan-Meier estimates give a good initial understanding of the survival distribution of the sample. However, in prepayment analysis one is additionally interested in comparing the impact on different characteristics among subjects or groups in the loan pool.

Suppose, a sample of loans of the size *j=1,..,n* contains again the survival time T_j for each individual *j*, each linked to a binary event indicator or depended variable δ_j.[14] Further assume that for each loan j, additional information of particular interest, referred to as *covariates* or *risk factors*, are captured in the vector $\boldsymbol{Z_j}(t) = [Z_{j1}(t), ..., Z_{jp}(t)]^t$,

[14] $\delta_j = 1$ event has occurred, and δ_j=0 if not.

(Klein and Moeschberger, 2005).[15] Note that risk factors might also simply be timely-independent $\boldsymbol{Z}$.

The most common model to relate the hazard rate (10) to a vector of covariates $\boldsymbol{Z}(t)$ is referred to as the Cox-Model (Cox, 1972):

$$h[t|\boldsymbol{Z}(t)] = h_0(t)exp[\boldsymbol{\beta Z}(t)] = h_0(t)exp[\sum_{k=1}^{p} \beta_k Z_k(t)] \tag{21}$$

where $h_0(t)$ is an arbitrary baseline hazard rate and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)$ is the vector of parameters. The Cox model is of semi-parametric nature because on the one hand it leaves the baseline hazard $h_0(t)$ unspecified (non-parametric). On the other hand, all covariate effects are treated parametrically. Therefore, covariates can interact in a simple linear fashion, as in a simple linear model.

Cox (1972) shows that $\boldsymbol{\beta}$ can be estimated without making any assumptions on the baseline hazard $h_0(t)$, using only the rank of failure and censored times. Suppose that $t_1<t_2<...<t_D$, as the D distinct, ordered event times. With respect to (21), the partial likelihood is then defined as:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{exp[\sum_{k=1}^{p} \beta_k Z_{(i)k}(t_i)]}{\sum_{j \in R(t_i)} exp[\sum_{k=1}^{p} \beta_k Z_{jk}} \tag{22}$$

Note that the denominator contains information of all loans who have not yet experienced an event (right-censored), whereas the numerator contains only information of loans experiencing an event. Cox's original approach in (22) assumes that only one single event can occur per time period. However, in most study setups, e.g. specifically in prepayment analysis, several subject face an event at the same time point (tied events). This issue was adressed by Efron (1977).[16] Let $R(t_i)$ be the set of numbers of subjects at risk just prior time t_i. Let d_i be the number of full prepayments at t_i and $\mathbb{D}_i$ the set of all individuals who prepay at t_i. Let $\boldsymbol{s}_i$ be the sum of vectors $\boldsymbol{Z_j}(t)$ over all loans prepaying at time t_i. With respect to (21), the partial likelihood is then defined as:

$$L_E(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{exp(\boldsymbol{\beta}^t \boldsymbol{s}_i)}{\prod_{j=1}^{d_i} [\sum_{k \in \mathbb{R}_i} exp(\boldsymbol{\beta}^t \boldsymbol{Z}_k) - \frac{j-1}{d_t} \sum_{k \in \mathbb{D}_i} exp(\boldsymbol{\beta}^t \boldsymbol{Z}_k)]} \tag{23}$$

$\boldsymbol{\beta}$ can be estimated using a maximum likelihood approach. Estimates can then be tested using univariate or multivariate Wald test (Klein and Moeschberger, 2005).[17]

[15] Or in the econometric terminology simply referred to as *regressors or independent variables.*

[16] There are actually more approaches to face tied event data. Since R-Software is using Efron's likelihood as default, only his concept will be amplified.

[17] With respect to the estimates of (23), local hypothesis of the Wald tests is $H_0 : \beta = 0$ and the global hypothesis $H_0 : \boldsymbol{\beta} = 0$.

The model is usually referred to as the Cox-Proportional Hazard model since it assumes the fundamental assumption of *proportionality* between the relative risk factors. For two subjects i and j, which differ in their covariates $Z_i(t)$ and $Z_j(t)$, their *relative hazard* can be defined as:

$$\frac{h[t|Z_i(t)]}{h[t|Z_j(t)]} = \frac{h_0(t)exp[\beta Z_i(t)]}{h_0(t)exp[\beta Z_j(t)]} = exp[(\beta(Z_i(t) - Z_j(t))] \tag{24}$$

Note that in (24), the relative hazard rate between $Z_i(t)$ and $Z_j(t)$ can be expressed by only one coefficient β, implying the key assumption of *proportionality* (Therneau and Grambsch, 2013).

At this point, it is crucial to understand that a violation of this assumption might lead to misleading and not reliable estimates, as we will see in the later. Therefore this assumption has to be tested. According Therneau and Grambsch (2013) the assumption can be evaluated by extending the Cox model (21) by time-dependent coefficients:

$$h[t|\boldsymbol{Z}(t)] = h_0(t)exp[\boldsymbol{\beta(t)Z}(t)] \tag{25}$$

Therneau and Grambsch (1994) show that if the estimated $\hat{\boldsymbol{\beta}}$ vary over time, e.g., the estimated coefficients for the yield-incentive slightly decrease over the age of the loan, the assumption of proportionality is violated. This can be specifically tested with help of the *Schoenfeld residuals*:

$$r_{ik} = Z_{ik} - E(Z_{ik}|R_i) \tag{26}$$

Equation (26) basically extends the usual residual concept to a right-censored data framework. It describes the difference between the observed value of covariate Z_ik and its expected value conditional on the risk set, i.e. the set of individuals who are still repaying at that time t (Stepanova and Thomas, 2002).

One can then test the assumption of proportionality, by simply plotting the Schoenfeld residuals (26) for each risk factor as a function of time. Then a regression line can be fit to detect time-trends. Whenever the line has a non-zero slope or is horizontal, this would indicate time-dependency of the estimated coefficients and hence a violation of proportional hazard. Clear graphical illustrations are given in section 3.3.2 Alternatively, one can test if the estimated slope of the regression line is $H_0 : \beta = 0$ using a univariate Wald test.

Lastly, to assess relative model fit, one can use the Akaike information criterion (AIC) (Klein and Moeschberger, 2005):

$$AIC = -2logL + 2p \tag{27}$$

where p is the number of regression parameters in the model and L is the likelihood function with respect to (23). A decrease in AIC states an increase in model fit, relative to another model.

3.2 Data

3.2.1 Study Design

The foundation of the empirical analysis is the *Fannie Mae Single-Family Historical Loan Performance Dataset*. This dataset is freely accessible on the online portal of the Fannie Mae and is updated quarterly.[18] The full database consists of 22 million single family, fixed-interest mortgage loans, starting in the year 2000. The data consists of monthly loan acquisitions of the Fannie Mae, containing individual loan information. To track loan records, a unique loan identifier is assigned to each loan. To specify the group of study, only first-time purchase loans with a term of 30 years were considered in the further analysis.

Survival studies have to set a clear time frame. It was decided to specify the starting period of the study from November 2009 to February 2010. During that time, the Fannie Mae acquired 77,240 first-time purchase, 30-year term, fixed-interest mortgage loans. This specific time frame was chosen to give the reader an overview of the current market environment. In particular, regulation and loan policies changed a lot specifically after the financial crisis of 2008 - 2009. Earlier study periods might hence be misleading since the current market reference is missing in general or it would incorporate with the abnormal market environment during the crisis. Further, we assume that at the beginning of 2010, the mortgage market entered normality. The end date was set to December 2015 due to a lack of regional mortgage-rates data, discussed in the later. Note that loans acquisitions happened after the starting period (delay-entry or left-truncation data) were not taken into account. In contrast to our analysis, some studies incorporate with these data (e.g., see the studies of Schultz (2015) or Deng et al. (2000)). However, research often argues that left-truncation might lead to a selection bias, referred to as *immortal-person time bias* (Rothman et al., 2008).

In case of a loan leaving the data set due to voluntary or involuntary prepayments, this event will be reported by a specified indicator. It delivers the essential information of how long the loan '*survived*'.[19] However, since involuntary prepayments due to credit-events are not in the scope of this thesis, only loans which faced a voluntary prepayment event and loans which had no event at the end of the study (right-censored loans) were taken into account. This assumption is easy to justify since the events of involuntary and voluntary prepayment are independent of each other: a voluntary prepaid loan cannot face an involuntary prepayment after the event has occurred and vice versa. It is important to note that these contracts do not contain prepayment penalties and

[18] http://www.fanniemae.com/portal/funding-the-market/data/loan-performance-data.html

[19] Indeed there were more indicators, like loan dispositions. These events were rare (around 150) and were not taken into account.

therefore prepayment can occur freely at every point in time. To reduce selection bias, additionally, loans in which the mortgage-rate was adjusted during the term were removed, as well as loans which were delinquent longer than 30 days in the last month of the study period. However, this 'clean-up' reduced the whole set only by around 1% (1092 loans).[20]

As mentioned in the theoretical framework, modern prepayment analysis incorporates data on the most geographically local level possible. In fact, thanks to the database of the *Federal Reserve Bank St. Louis*, macroeconomic data is provided on a regional or MSA (metropolitan statistical area) level which can easily be merged to the loan-level dataset since it includes all necessary geographical information. However precise, granular data comes usually with fewer observations. In general, out of the initial 77,240 loan observations, only 31,001 loans could be matched to their granular macro counterparts. An overview of the geographical distributions of the loans is given in the Appendix.

Note that to operate with time-dependent covariates, the data set had to be transformed into a panel structure, referred to as *episode spliting* (Therneau, Crowson, et al., 2017). Depending on the individual survival time, every single loan now consists of multiple observations. In general, the data set consists of 1,186,146 observations after the panel transformation. Note that time-fixed covariates, e.g., credit-volume or geographical information, remain constant over each observation.

3.2.2 Descriptive Statistics

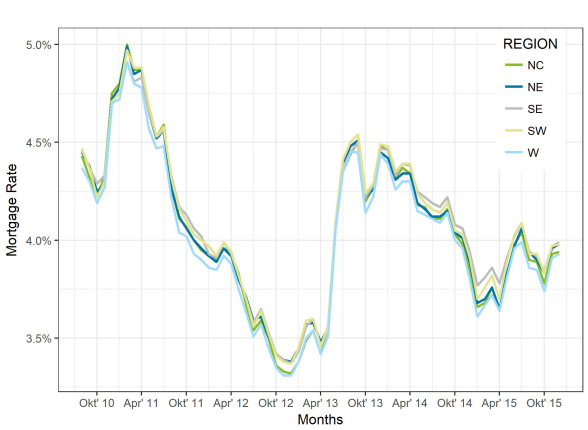
In this chapter the reader will be provided with a short overview of the analyzed risk factors or covariates, considered in the following Cox regression in section 3.3.2. The set of risk factors was chosen with respect to the economic discussion in 2.4.1.

First, to incorporate for the yield incentive, average mortgage interest rates for the five geographical regions were considered: Northeast, Northcentral, Southeast, Southwest, and West, starting from January 2010 to December 2015. The yield incentive was then simply modeled as the spread between the averaged mortgage rate $MR_k(t)$, of region k and the contractually agreed fixed note rate NR_i of the individual loan i :

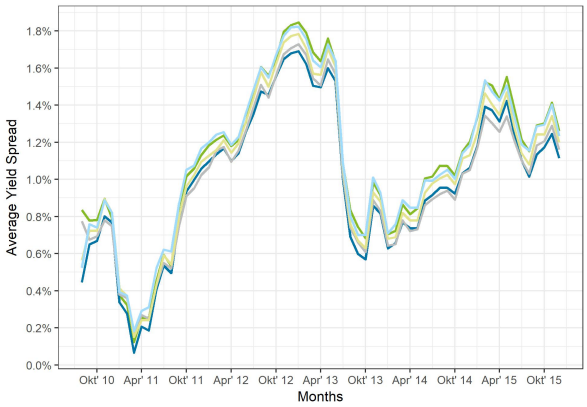
$$YSPREAD_i(t) = MR_k(t) - NR_i \tag{28}$$

A positive yield spread with respect to (28) indicates that the value is in-the-money, while negative values indicate that the loan is out-of-money, from a pure yield perspective. Figure 3 illustrates the regional, average mortgage rates itself, as well as the

[20]Furtheremore, it was additionally checked if any remaining loans were part of the governmental *Home Affordable Refinance Program* (HARP) but this was not the case.



Average Mortgage Rates among Regions



Aggregated Average Yield Spreads among Regions

Figure 3: Yield Incentive

average yield spreads aggregated over all individual loans and regions. As it can be seen, they do not diverge significantly among the regions. However due to the major importance of this factor, even slight differences might influence modeling outcome and hence differences have to be incorporated. Unfortunately, data for the average, regional mortgage rate does not exist for further dates after December 2015. Nonetheless, it was decided to rather take a reduced period of study into account, then a loss in precision.

As elaborated in chapter 2.4, house-price development is assumed to be 2nd major driver of refinancing decisions, specifically on an individual LTV level. Dynamic LTVs for each loan i at month t can simply be computed by:

$$LTV_i(t) = \frac{OBAL_i(t)}{PROP_i(t)} \tag{29}$$

where $OBAL$ is the monthly remaining outstanding balance and PROP is the value of the underlying property of loan i.

Individual outstanding balances of each month for each loan are provided by the Fannie Mae data set, as well as the initial LTV, from which the initial house-price for each loan at t=0 can then extracted.

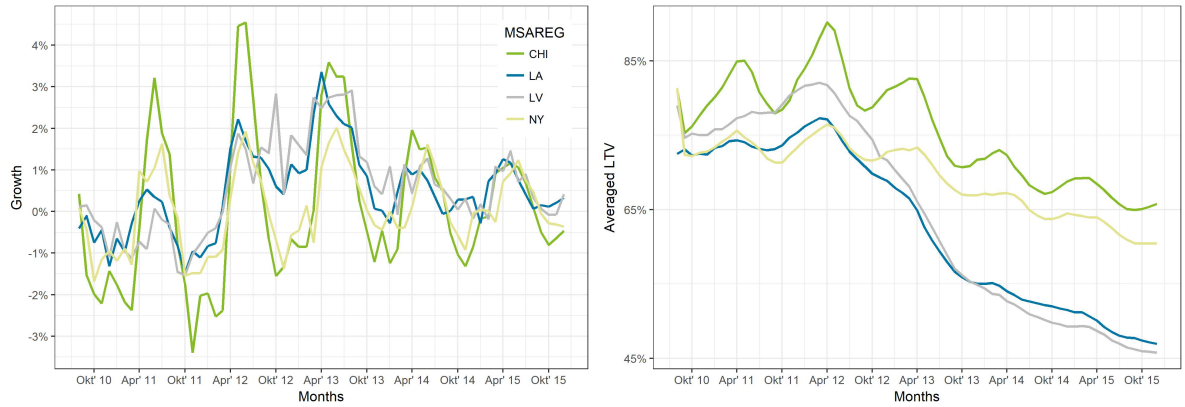
To remodel further house-price development, monthly growth rates were extracted from the leading S&P / Case-Shiller Home Price Indices, which is provided for 20 selected MSAs.[21]. Each MSA index measures the average change of the market value of a single-family house using the repeat-sales method. This method simply computes observable sales prices deviations of the mortgages sold over a specified term in the specified geographical area (Dow Jones Indices, 2017).

[21]The list of the 20 specific MSAs can be found in the Appendix.

The property development of a loan i can then simply be remodeled by:

$$PROP_t^{ik} = PROP_{t-1}^{ik}(1 + g_t^k) \tag{30}$$

where g is the associated, extracted Case-Shiller growth rate at time t of MSA k.



Averaged Case-Shiller Index Growth

Averaged LTVs

Figure 4: Dynamic LTV

Figure 4 shows exemplary Shiller-Growth rates and averaged LTVs ratios for five chosen MSA. These plots aim to highlight the importance of local, granular data. For instance, Chicago's house-price development seems quite noisy with strong amplification in both directions. This presence is felt explicitly on the LTVs in the right plot. These effects would not be observable with national data.

Note that the definition of dynamic LTVs (29) does not only describe property appreciation because dynamic LTVs with respect to 29 additionally measure remaining outstanding balance or in more general terms, the subjects personal debt-ratio at each month is taken into account. Thus, this variable should rather be regarded as a measure of the subject's financial standing. According to Deng et al. (2000) is also referred as the individual's *equity* ratio. Also, it should be stated that dynamic LTVs implicitly contain a negative time-trend since outstanding balance is shrinking due to scheduled repayments.

Furthermore, with respect to chapter 2.4, to following variables particular attention was paid:[22]

- Initial Credit Volume (VOL) in $10,000, timely-fixed
- Initial Debt-to-Income Ratio (DTI) in %, timely-fixed
- Initial Credit Score (FICO), timely-fixed
- Local unemployment Rate (UNEMP) in %, monthly on MSA level

[22]The shortcut in the brackets reports the names of the variable of the data set.

- Indicator of a Refinance Tax of the State New York (TAX), dummy variable

DTI and credit scores were considered in the analysis, but indeed their impact should not be overinterpreted. Unfortunately, their effects are only of limited value since values are reported just at the beginning of the tracking period and are not updated during the loan's lifetime. Also, it was of further interest to incorporate for general debt-aversion effects. Therefore the initial credit volumes of every single subject were taken into account. Besides this, additional focus laid of the impact on local macroeconomic factors. Consequently, local unemployment rates of the local MSA, as well as the refinancing tax of the state New York were considered in the empirical analysis.

A complete summary statics of the variables can be found in Table 2.

Table 2: Summary Statistic

Covariates*	Mean	St. Dev.	Min	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	Max
YSPREAD	1.013	0.557	−1.125	0.615	1.000	1.400	3.755
LTV	69.401	16.105	0	60	73	81	123
VOL	23.783	14.639	1.000	12.400	20.500	32.000	112.900
FICO	760.110	39.924	580	736	770	791	830
DTI	36.618	11.292	1	29	37	44	64
UNEMP	7.928	2.009	3.300	6.200	8.000	9.200	14.000
TAX.NUM	0.201	0.401	0	0	0	0	1

* Total Number of Loans at t_0 N=30,001. And 1,186,146 observations of the whole period of study. Note that YPSREAD, LTV and UNEMP are averaged over time.

[1] in $ 10,000.

[2] The Tax mean indicates that 20.1% of the loans are located in the state of New York.

Again note that the goal of this thesis is to provide only statistical inference. Finding the best model fit is not within the scope of this analysis, and therefore additional interactions between the described variables were not considered.

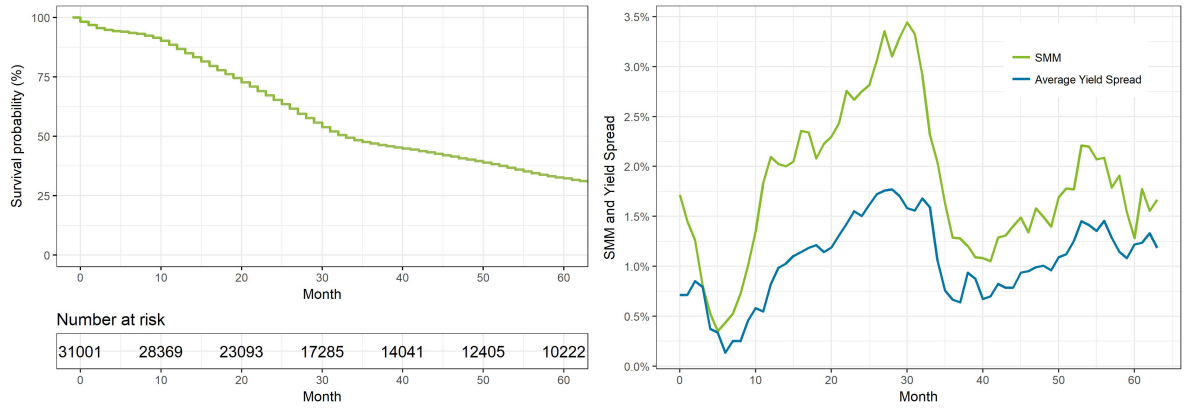
3.3 Results

3.3.1 Empirical Prepayment Rates

As a starting point, it is helpful to give the reader a general overview of the survival distribution of the full sample. Due to its high illustrative power, a Kaplan-Meier

estimation with respect to (17) and (18) is conducted on the full sample over the defined period of study without further specification. This unspecified, non-parametric model is usually referred to as the null model (Schultz, 2015).

The left plot in Figure 5 illustrates the estimation of the survival function over the specified term of 65 months. Due to the large number of observations, confidence bounds are extremely narrow in this estimation and almost not recognizable. The illustration additionally contains the risk table which provides statistics on the total number of loans at each point in time. In general, out of the 31,001 loans in the beginning, only 10,222 remained in month 65, which indicates that around 30% of all loans fully repaid until the end of the study period. The hazard rates or, mathematically speaking, the slope of the estimated survival function is steeper specifically around month 10 and flattens again around month 35.



Survival Curve | Empirical $\widehat{SMM}$

Figure 5: Kaplan-Meier Estimates: Null Model

The right plot in Figure 5 illustrates observed empirical hazard rate $\widehat{SMM}$ of the pool. Additionally the average yield spread of Figure 3 was added to demonstrates the strong graphical correlation between the mortgage rates and the empirical prepayment rates. Unmistakable, prepayment rates grow and shrink mostly in the same months. It increases whenever the yield spread rises and vice versa. Note that yield incentive was modeled ex-post as a one lag variable. Interestingly, one lag seems to be enough to aptly describe prepayment behavior because both time series appear not to move staggered.[23]

As mentioned in the methodological framework of this chapter 3.1.1, some authors like Deng et al. (2000) or Schultz (2015) claim that the estimated empirical $\widehat{SMM}$ in (19) are equal to the realized cashflow SMM with respect to (4). To avoid further misconception, this claim will be analyzed in more detail by comparing both measures.

[23]Note that this is only a graphical claim.

Cashflow prepayment rates can be extracted using an amortization calculation[24] on the full pool of loans with respect to the real monthly prepayment cashflows. Cashflow and empirical $\widehat{SMM}$ can then simply be plotted against each other. Figure 6 illustrates these findings.

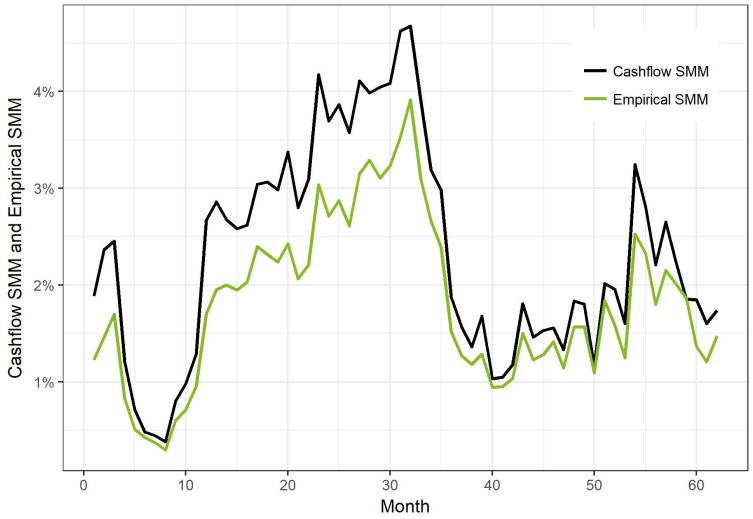


Figure 6: Cashflow SMM vs. Empirical $\widehat{SMM}$

First of all, both series follow the same pattern over the whole period of study. However, the $\widehat{SMM}$ underestimates the realized cash flow up to around 0.8%. This comes from the obvious fact that Kaplan-Meier estimates (or survival analysis in general) only reflect full prepayment events since empirical prepayment rates are modeled as a counting process. Prepayment cashflows, which do not lead to a reduction in observable numbers are therefore not captured by construction. Most significant differences are explicitly found between month 10 and 20, where yield spreads grow due to a sharp decline in mortgage rates, as it was observed in Figure 3 and Figure 5. During that period, not only the number of full repayments rose drastically, but also the amount of prepayment cash flows itself. Nonetheless, the $\widehat{SMM}$ reflects a reasonable estimate since full prepayments make up most of prepayment cash flows, thanks to the US market environment.

If these circumstances are given, the described Kaplan-Meier approach can hence be understood as one way to meet parts of the new EBA requirements on the IRRBB regarding behavioral options.

However, note that this indeed does not hold in a German private mortgage environment, where cancellation or full repayments is highly restricted. In the first ten years, prepayment cannot be modeled using a counting process and specifically Cox-Regression is not applicable. Nevertheless, for uncovered retail loans, this approach

[24] As described in chapter 2.3.1 in equation (1) - (3). The whole analysis can be found in the R-Script *SMM_Compare.R* attached in the folder.

might even be useful in a German market environment

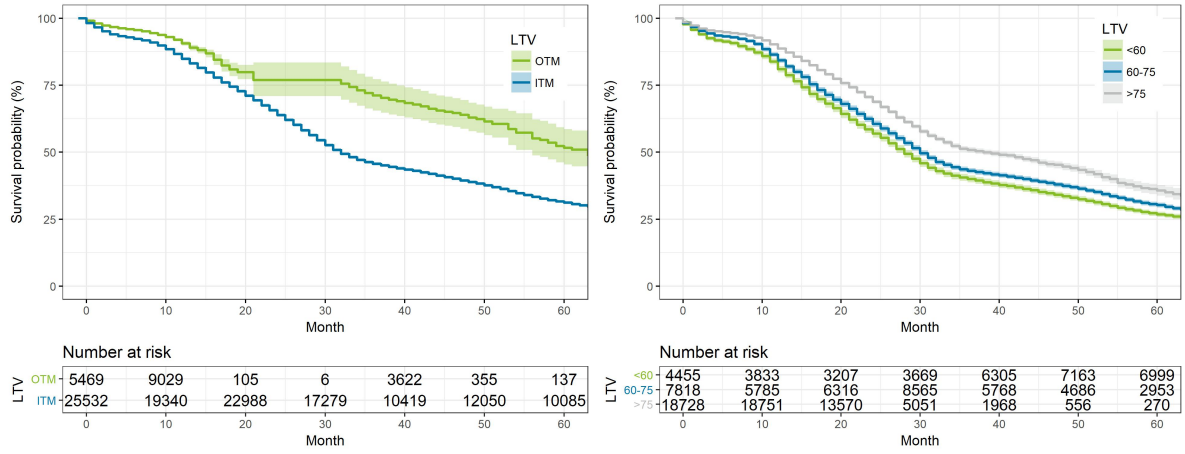
3.3.2 Cox-Regression Estimates

Next, the reader will get an overview of the estimated results.[25] However, before continuing the empirical analysis, the analyzed covariates have to be simplified first, since the Cox-Method is not designed to incorporate large amounts of timely-varying continuous data.[26] This is a major issue specifically for the two most major prepayment drivers: yield incentive and LTV. One way to handle this issue is by simple categorization.

Firstly, the yield incentive variable YSPREAD(t) was grouped, with respect to (28), in following fashion:

$$YSPREAD(t) = \begin{cases} \text{OTM}, & \text{if YSPREAD(t)} < 0.5 \\ \text{ITM}, & \text{if YSPREAD(t)} \geq 0.5 \end{cases}$$

Observations with a yield spread below 0.5% were labeled as out-of-money (OTM), whereas observations greater then 0.5% are marked as in-the-money observations (ITM). The threshold of 0.5% was set to incorporate the behavioral and economical aspect of private loan owners. First, transaction costs have to be considered. Refinancing usually lead to banking fees and documentation cost (Joshi et al., 2016). Secondly, irrationality aspect of private homeowners has to be acknowledged. One can assume that homeowners would not refinance in a marginal positive yield incentive scenario.[27]



Yield Incentive Dynamic LTVs

Figure 7: Survival Curves: Yield Incentive & Dynamic LTV

[25] All results can be found in the R-Script *Regressions.R*.

[26] Note that at this point, the term continuous and discrete is referred to data structure and should not be mistaken by its mathematical definition.

[27] A precise estimation of the threshold could be a topic of further research.

The left plot in Figure 7 shows the Kaplan-Meier estimates for the survival function of the categorized yield spread. Both groups can be separated because survival time of ITM is apparently higher. However, due to the downward sloping yield regime in the period of study, the number of OTM observations is distinctly smaller. Specifically between month 20 and 30, when mortgages dropped and yield spreads rose substantial as illustrated before, only a fraction of loans could be labeled as OTM. Indeed at some points in time, none of the loans could be marked as OTM. In that case, it is assumed that the survival rates are equal to the last available observation. Confidence bounds are thus wide. Nonetheless, the difference between the groups is unmistakable.

Similar to the model of Deng et al. (2000), at every time-point, observations are grouped according to the computed, dynamic LTV with respect to (28). Note that loans might switch their group.

$$LTV(t) = \begin{cases} LTV(t) < 60 \\ 60 \geq LTV(t) < 75 \\ LTV(t) \geq 75 \end{cases}$$

The right Figure of 7 illustrates the Kaplan-Meier estimates of the three subgroups. Survival distributions are distinguishable in all three cases. Low LTV observations (LTV<60) show the lowest rate of survival, whereas the survival probability increases with increasing LTV. These findings indicate a linear functional relationship between dynamic LTV and survival rates since the survival functions do not cross among the three groups. In general, results are in line with theory since low LTV ratios indicate either an increase in property value leading, for example, to a cashout refinance or indicate a substantial decrease in the outstanding balance. Also note that in group LTV>75, the number of observations shrinks, leading to an increase in confidence bounds. This mainly results from the fact that on average LTV ratios have slightly declining linear relationship with time, resulting from a general rise in scheduled principal payments with respect to the annuity payments[28] over time. Therefore mainly loans with a relative high outstanding balance due to low prepayments and (or) a weak increase in property value will be assigned to this group at the end of the study period.

Table 3 presents the results of the full model estimation including the categorized major variables, as well as further personal and macro economic characteristics, as described in chapter 3.2.2. First of all, all estimates show narrow confidence bounds. The H_0 that the coefficient is equal to 0 can be rejected for each estimate. Also, the global Wald statistic can be rejected. At this point, it is important to mention how coefficients can be interpreted. First of all, note that all coefficients are reported in their exponential form. In general, an estimate of $exp(\hat{\beta}) > 1$ indicates an increase in relative risk, whereas $exp(\hat{\beta}) < 1$ indicates a negative impact on negative risk.

[28] As defined in 1 in section 2.3.1

Table 3: Estimates of Model I

	$exp(\hat{b})$
SPREAD.ITM	1.801*** (1.740, 1.861)
LTV.<60[1]	1.135*** (1.099, 1.170)
LTV.>75[1]	0.815*** (0.780, 0.850)
VOL	1.028*** (1.028, 1.029)
DTI	0.993*** (0.992, 0.994)
FICO	1.003*** (1.002, 1.003)
UNEMP	0.962*** (0.952, 0.971)
TAX.NY	0.697*** (0.660, 0.733)
Global Wald Test	6,152.100*** (df = 8)
AIC	414,248

Note: *p<0.1; **p<0.05; ***p<0.01
The brackets report the 95%- Confidence Intervals
[1] baseline group: 65<LTV<75

Furthermore, categorical estimates have to be interpreted with respect to their relative counterparts. For instance, an estimate of 1.801 for the $SPREAD.ITM$ covariate states that an ATM loan is by 80.01% more likely to prepay then an OTM loan if the proportionality assumption holds. Also, continuous variables have to be interpreted relatively. For instance, an estimate of $\exp(\widehat{DTI})$=0.993 states that a loan with a DTI ratio of, e.g., 61% would be (1-0.993)=0.07% less likely to prepay in the monitored period than a loan with a DTI of 60% and vice versa. This applies to all relative distances of this variable. A loan with a reported DTI of 60% is by (1- $(0.997)^{20}$)=5.83% less likely to prepay then a loan with a DTI of 40%. However again, results are only reliable under the condition of proportionality. Before conducting economic interpretation of the results, hence first the assumption has to be tested using a graphical and a statistical test as proposed by Therneau and Grambsch (2013) with respect to (25).

Figure 8 shows the graphical illustration of the Schoenfeld residual tests for four specific estimates. In general, the test estimates the coefficients as a function of time. The left-corner plot illustrates the estimated coefficient for the ITM yield spread category. It starts with a value of around exp(1.1)=3.00 and then shrinks to an estimate of around exp(0.5)=1.649, relative to OTM loans. Therefore a time-dependency exists implying a violation of the proportionality assumption. The time-averaged estimated coefficient of 1.801 as reported in Table 3 is therefore misleading.

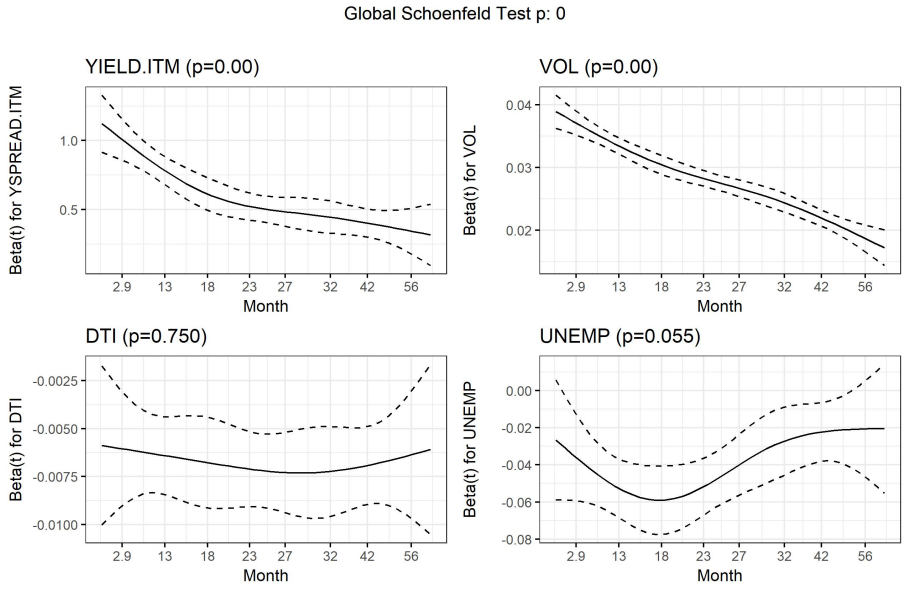


Figure 8: Graphical Proportionality Test: Model I

One should not be surprised by the result since it illustrates the burn-out effect mentioned in chapter 2.4.1. Due to adverse selection, only poor performing loans remain in the pool, leading to less sensitive reaction to yield incentive with increasing time. Note that the Schoenfeld test does neither hold for the VOL, where one can find a clear linear time trend. Exemplary, two timely independent coefficients DTI and TAX.NY are additionally illustrated. In both cases, they fairly remain horizontal indicating a non-violation of the proportional hazard assumption.

Table 4: Proportionality Test of Model I

	p-values
YSPREAD.ITM	0.000
LTV.<60	0.000
LTV.>80	0.000
VOL	0.000
DTI	0.750
FICO	0.000
UNEMP	0.055
TAX.NY	0.953
Global Wald Test	0.000 (df = 8)

Table 5: Estimates of Model II

	$exp(\hat{b})$
YSPREAD.ITM	2.222*** (2.133, 2.312)
LTV.<60[1]	1.360*** (1.286, 1.434)
LTV.>75[1]	0.714*** (0.645, 0.783)
DTI	0.994*** (0.992, 0.995)
FICO	1.005*** (1.005, 1.006)
UNEMP	0.961*** (0.951, 0.970)
VOL	1.038*** (1.036, 1.039)
TAX.NY	0.680*** (0.643, 0.717)
t*YSPREAD.ITM	0.989*** (0.986, 0.993)
t*LTV.<60[1]	0.994*** (0.992, 0.996)
t*LTV.>75[1]	1.009*** (1.006, 1.012)
t*FICO	0.999*** (0.999, 1.000)
t*VOL	0.999*** (0.999, 1.000)
Global Wald Test	6,603.400*** (df = 13)
AIC	413,842

Note: *p<0.1; **p<0.05; ***p<0.01
The brackets report the 95%- Confidence Intervals
[1] baseline group: 65<LTV<75

Additionally, the results of the proportionality test statics are given in Table 4. It tests the H_0 whether the slope of the respective time-coefficient function concerning the Figures in 8 are statistically equal to 0. A rejection of the test statistic, therefore, implies a violation of the PH assumption. As it gets clear, low p-values indicate that except the variables DTI and the two macro factors UNEMP and TAX.NY, the premises are profoundly violated. Also the global test speaks for a substantial violation of the PH assumption of the full model and the interpretation of the coefficients should be treated with caution.

Nonetheless, there are different ways to handle non-proportional covariates. As proposed by Fox (2002) and Therneau, Crowson, et al. (2017) one can incorporate linear time-trends by simply building time-interactions into the model. The results of the modified model are given in Table 5. As you can see again, all results are strongly significant. Specifically, the strong significance of the time-effects indicates a time-dependency between the problematic variables. Note that, a further reduction of the

AIC parameter with respect to (27) from 414,248 to 413,842 states an improved model fit and disproves over-fitting.

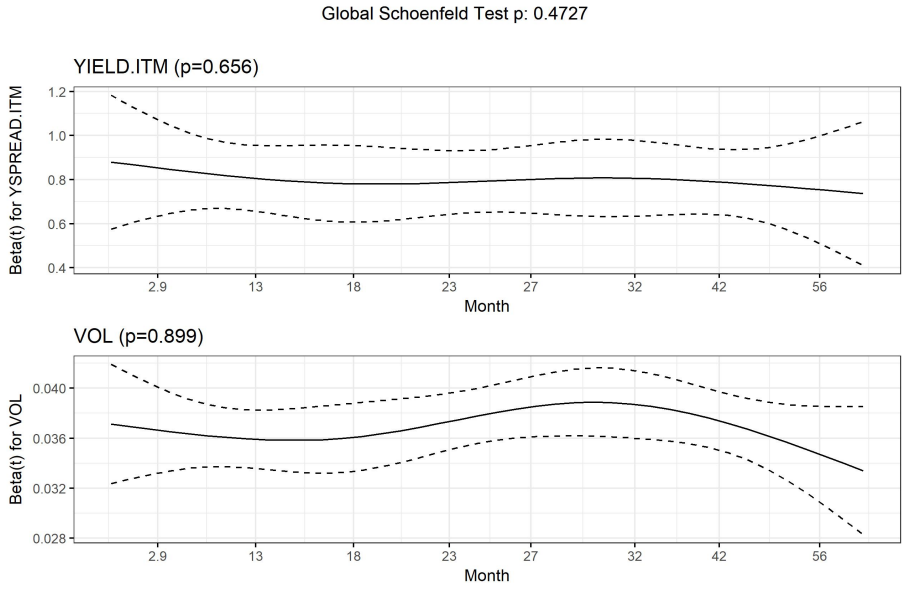


Figure 9: Graphical Proportionality Test: Model II

	p-values
YSPREAD.ITM	0.656
LTV.<60	0.263
LTV.>80	0.502
VOL	0.899
DTI	0.976
FICO	0.024
UNEMP	0.059
TAX.NY	0.785
Global Wald Test	0.4727 (df = 13)

Table 6: Proportionality Test for Modell II

As you can see in Figure 9 and Table 6 proportional hazard assumptions hold graphically and statistically for all covariates on a 5% significance level (except for the credit score factor FICO). Also, the global Wald test cannot be rejected. Therefore, time-effects have successfully be captured and parametrized by the additional variables. The estimates for the credit risk variable seems to follow a statistical significant non-linear time trend, and therefore time-interactions terms cannot capture these effects. Thus the estimate is therefore not reliable, and interpretation should only be made with

caution.

In general, all estimates are thereby in line with the economic theory described in 2.4. The estimate of the yield incentive factor states that over the whole period of study the effect of being *in-the-money* increases the probability of prepaying by 122.22%, relatively to loans being out-of-money. This is a reasonable estimate and indeed indicates the most substantial effect of all covariates. Also, as Schoenfeld tests suggest, there is again a clear proof of burnout, since there exists a strong significant down-ward sloping linear time-trend. Indeed, the time interaction covariate parameterized this effect. It states that the relative positive effect of being ITM vs. OTM is shrinking with a rate of 1.1% per month.

Also, the results for LTV are in line with economic theory. Note that the relative baseline group was set to observations labeled as LTVs of group 65<LTV<75. First, again a linear functional relationship exists. Being in the low LTV<60 group increases the probability of prepaying by 36.10%, but its effect decreases with time by approximately 0.6% per month. On the other hand, an increase in LTV reduces the probability of prepaying by 28.6% compared to the baseline group while hazard increases with time by 0.9%. It can hence be concluded that low LTV or high equity loans prepay the fastest (e.g., to cashout refinance), whereas low equity loans tend to prepay slower due to several reasons discussed in 2.4.1.

These positive and negative time trends of the linear interaction estimates for the LTV groups prove another form of selection bias in the mortgage market. High equity observations leave the pool of loans relatively faster in the beginning since refinancing opportunities are more accessible compared to lower LTV groups. On the other hand, the positive time-trend of the low equity loans indicates that only with approaching time, refinancing opportunities can be realized, e.g., due to an increase in property value, resulting in an increase in hazard over time.

Other control-variables also are in line with theory. An increase in credit volume by \$ 10,000 positively impacts the hazard of prepaying by 3.80%. This is reasonable since high-volume loan subjects have more access to refinancing alternatives. It might also be regarded as a reflection of the debt aversion of loan borrowers. Indeed this effect is relatively strong. It implies that for instance a loan of \$500,00 is 2,54-times[29] more likely to prepay then a medium-size loan of \$250,000. Credite volume effects however slightly but significantly decrease linearly by 0.01% per month.

Not surprisingly, increasing initial DTI estimates prove a negative impact on prepayment hazard by 0.06%. This is reasonable since high debt subjects struggle in finding general refinancing opportunity. As reported before, a high-debt loan with an initial

[29] $1.038^{20} = 2.54$.

DTI of e.g., 60% is by (1- $(0.997)^{20}$)=5.83% less likely to prepay then a loan with a DTI of 40%. FICO seems to have a positive impact. An increase in FICO score by one point increases the probability of prepaying by 0.05% on average. However, FICO estimates should be interpreted with caution since proportionality can only be rejected at the 2% confidence-level.

Also, macroeconomic covariates prove the theoretical findings and underline the importance of granular local data. Indeed an increase in unemployment rate in the MSA reduces prepayment by 3.9% on average, again due to less refinancing opportunities, as assumed correctly in chapter 2.4.1 . Not surprisingly, the refinancing tax of the state of New York impacts prepayments negatively by 32%, compared to the national average. Local economic conditions have a clear, distinctive effect on rates and therefore require particular attention.

4 Conclusion

To introduce the topic on prepayment to the reader, first certain theoretical and regulatory aspects were provided. It was found that the definition of prepayment does not only depend on the instrument's legal entity. This is of special importance when it comes to identification and measurement. Also, clear new requirements are set by the EBA (2017). Prepayment analysis is hence not only mandatory to deepen customer-understanding, but also has to be taken into account from a regulatory perspective.

In a mortgage framework, risk can be captured by the SMM (or CPR) which under certain circumstances might profoundly impact the economic value of the underlying instrument. Moreover, we have seen that prepayment rates should be regarded at least as a function of the interest-rate itself. However, in a mortgage framework, prepayment rates itself have to take further dimensions into account. House-price appreciation, personal financial status or local macroeconomic environment play the most significant role in driving prepayments.

The empirical study on the Fannie Mae Data concretely presented the theoretical aspects and resulting regulatory requirements. Therefore, the Kaplan-Meier estimator was introduced as a simple but powerful approach to estimate empirical prepayment rates. Also, Cox regression was used to provide statistical inference from which behavioral assumptions were confirmed. It was shown that customer behavior in the US retail market is highly complex. A pure yield incentive can be regarded as the major driver. However, it was found that prepayment decision is also driven by the personal financial situation of the subject and its local economic environment. Note that particular attention should be paid on the assumption of proportionality when applying

Cox-Regression.

However, during the empirical research, limitations of survival analysis were revealed. Empirical $\widehat{SMM}$ rates using a Kaplan-Meier estimation might underestimate the prepayment risk since only full repayments in the loan pool are taken into account. In particular this makes the method only applicable if prepayment behavior is not restricted. It can hence not be applied to similar German mortgage loans due to the unique legal boundaries. Nonetheless, survival analysis could be of particular interest for German institutes which target less-restricted private loan segments like retail loan markets. Therefore, it could be used as a useful tool to fulfill possible regulatory requirements and at the same time deepening customer understanding. The empirical analysis conducted could, therefore, be regarded as a practical use case from both perspectives.

Note that another limitation of survival analysis is its weak predictive power as stated by Schultz (2015). Indeed modern machine learning techniques should preferably be applied for predictive purposes what could be another interesting topic for upcoming research in the field of prepayments. Sirignano et al. (2017) made the first contribution in this context. Despite its limitations, survival analysis can still be regarded as an efficient way to measure risk and can be understood as a preliminary analysis for further research.

References

Banerjee, Steve, Anand Bhattacharya, and BIll Berliner (2016). "Contemporary Challenges in Loan-Level Prepayment Modeling". In: *The Handbook of Mortgage-Backed Securities*. Ed. by Frank J. Fabozzi. Oxford University Press. Chap. 26, pp. 561–581.

BCBS (2016). *Interest Rate Risk in the Banking Book*. Bank for International Settlements.

Brennan, Michael J. and Eduardo S. Schwartz (1977). "Savings Bonds, Retractable Bonds and Callable Bonds". In: *Journal of Financial Economics* 5(1), pp. 67–88.

Choudhry, Moorad (2010). *Fixed-income Securities and Derivatives Handbook*. Bloomberg. 2nd. John Wiley & Sons. Chap. 11, p. 207.

Cox, David Roxbee (1972). "Regression Models and Life-Tables". In: *Journal of the Royal Statistical Society*. Series B (Methodological) Vol. 34(No. 2.) pp. 187 - 220.

Deng, Yongheng, John M Quigley, and Robert Order (2000). "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options". In: *Econometrica* 68(2), pp. 275–307.

Dow Jones Indices, S&P (2017). *S&P CoreLogic Case-Shiller Home Price Indices Methodology*. S&P Global.

EBA (Oct. 31, 2017). *Draft Guidelines on the Management of Interest Rate Risk arising from non-trading Book Activities*. EBA/CP/2017/19. European Banking Authority.

Efron, Bradley (1977). "The Efficiency of Cox's Likelihood function for Censored Data". In: *Journal of the American statistical Association* 72(359), pp. 557–565.

Fabozzi, Frank J. (2016). "Cash Flow Mathematics for Agency Mortgage-Backed Securities". In: *The Handbook of Mortgage-Backed Securities*. Ed. by Frank J. Fabbozzi. 7th. Oxford University Press. Chap. 3, pp. 87–104.

Fabozzi, Frank J., Gerald W. Buetow, and Robert R. Johnson (2005). "Measuring Interest-Rate Risk". In: *The Handbook of Fixed-Income Securities*. Chap. 9, pp. 183–228.

Fabozzi, Frank J., Glenn Schultz, and Bill McCoy (2016). "Agency Mortgage Passthrough Securities". In: *The Handbook of Mortgage-Backed Securities*. Ed. by Frank J. Fabbozzi. 7th. Oxford University Press. Chap. 6, pp. 167–194.

Fox, John (2002). "Cox Proportional-Hazards Regression for Survival Data". In: *An R and S-PLUS companion to applied regression* 2002.

Hull, John and Alan White (1990a). "Pricing Interest rate Derivative Securities". In: *The Review of Financial Studies* 3(4), pp. 573–592.

Hull, John and Alan White (1990b). "Valuing derivative securities using the explicit finite difference method". In: *Journal of Financial and Quantitative Analysis* 25(1), pp. 87–100.

Joshi, Rajashri, Tom Davis, and Bill McCoy (2016). "Valuation of Mortgage-Backed Securities". In: *The Handbook of Mortgage-Backed Securities*. Ed. by Frank J. Fabozzi. Oxford University Press. Chap. 24, pp. 503–531.

Klein, John P and Melvin L Moeschberger (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media.

Rothman, Kenneth J, Sander Greenland, Timothy L Lash, et al. (2008). "Cohort Studies". In: *Modern epidemiology*. 3rd. Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia. Chap. 6, pp. 100–110.

Schultz, Glenn (2015). *Investing in Mortgage-Backed and Asset-Backed Securities*. 1st. John Wiley & Sons, Inc.

Schwartz, Eduardo S. and Walter N. Torous (1989). "Prepayment and the Valuation of Mortgage-Backed Securities". In: *The Journal of Finance* 44(2), pp. 375–392.

Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke (2017). "Deep Learning for Mortgage Risk". In: *Working Paper Stanford University*.

Stanton, Richard (1995). "Rational Prepayment and the Valuation Mortgage-Backed Securities". In: *The Review of financial studies* 8(3), pp. 677–708.

Stepanova, Maria and Lyn Thomas (2002). "Survival Analysis Methods for Personal Loan Data". In: *UBS AG: Operations Research* 50(2), pp. 277–289.

Therneau, Terry, Cindy Crowson, and Elizabeth Atkinson (2017). "Using time dependent Covariates and time dependent Coefficients in the Cox Model". In: *Survival Vignettes*.

Therneau, Terry and Patricia Grambsch (1994). "Proportional Hazards Tests and Diagnostics based on Weighted Residuals". In: *Biometrika* 81(3), pp. 515–526.

Therneau, Terry and Patricia Grambsch (2013). *Modeling Survival Data: Extending the Cox model*. Springer Science & Business Media.

Weiner, Jonathon (2016). "Modeling Prepayments and Defaults for MBS Valuation". In: *The Handbook of Mortgage-Backed Securities*. Ed. by Frank J. Fabozzi. Oxford University Press. Chap. 25, pp. 531–560.

5 Appendix

Month	Should be Payment	Principal	Prepayment(SMM)	Interest	Full Payments	Balance
0						$1.000.000,00
1	$4.216,04	$1.716,04	$1.679,26	$2.500,00	$5.895,30	$996.604,70
2	$4.208,95	$1.717,44	$1.673,54	$2.491,51	$5.882,49	$993.213,72
3	$4.201,87	$1.718,83	$1.667,84	$2.483,03	$5.869,70	$989.827,06
4	$4.194,80	$1.720,23	$1.662,14	$2.474,57	$5.856,94	$986.444,69
5	$4.187,74	$1.721,63	$1.656,44	$2.466,11	$5.844,19	$983.066,61
6	$4.180,70	$1.723,03	$1.650,76	$2.457,67	$5.831,46	$979.692,82
7	$4.173,67	$1.724,43	$1.645,08	$2.449,23	$5.818,75	$976.323,30
8	$4.166,65	$1.725,84	$1.639,41	$2.440,81	$5.806,06	$972.958,05
9	$4.159,64	$1.727,24	$1.633,75	$2.432,40	$5.793,39	$969.597,06
10	$4.152,64	$1.728,65	$1.628,09	$2.423,99	$5.780,73	$966.240,32
11	$4.145,65	$1.730,05	$1.622,44	$2.415,60	$5.768,10	$962.887,82
12	$4.138,68	$1.731,46	$1.616,80	$2.407,22	$5.755,48	$959.539,56
13	$4.131,72	$1.732,87	$1.611,17	$2.398,85	$5.742,89	$956.195,52
14	$4.124,77	$1.734,28	$1.605,54	$2.390,49	$5.730,31	$952.855,70
15	$4.117,83	$1.735,69	$1.599,92	$2.382,14	$5.717,75	$949.520,09
16	$4.110,90	$1.737,10	$1.594,31	$2.373,80	$5.705,21	$946.188,68
17	$4.103,99	$1.738,52	$1.588,70	$2.365,47	$5.692,69	$942.861,46
18	$4.097,09	$1.739,93	$1.583,10	$2.357,15	$5.680,19	$939.538,43
19	$4.090,19	$1.741,35	$1.577,51	$2.348,85	$5.667,70	$936.219,57
20	$4.083,31	$1.742,76	$1.571,92	$2.340,55	$5.655,24	$932.904,88
342	$2.374,54	$2.264,52	$70,22	$110,02	$2.444,76	$41.673,05
343	$2.370,54	$2.266,36	$66,29	$104,18	$2.436,83	$39.340,40
344	$2.366,56	$2.268,20	$62,36	$98,35	$2.428,92	$37.009,83
345	$2.362,57	$2.270,05	$58,44	$92,52	$2.421,01	$34.681,34
346	$2.358,60	$2.271,90	$54,52	$86,70	$2.413,12	$32.354,93
347	$2.354,63	$2.273,75	$50,60	$80,89	$2.405,23	$30.030,58
348	$2.350,67	$2.275,60	$46,69	$75,08	$2.397,36	$27.708,30
349	$2.346,72	$2.277,45	$42,78	$69,27	$2.389,50	$25.388,07
350	$2.342,77	$2.279,30	$38,87	$63,47	$2.381,64	$23.069,90
351	$2.338,83	$2.281,16	$34,97	$57,67	$2.373,80	$20.753,77
352	$2.334,90	$2.283,01	$31,07	$51,88	$2.365,97	$18.439,69
353	$2.330,97	$2.284,87	$27,17	$46,10	$2.358,14	$16.127,65
354	$2.327,05	$2.286,73	$23,28	$40,32	$2.350,33	$13.817,64
355	$2.323,13	$2.288,59	$19,39	$34,54	$2.342,53	$11.509,66
356	$2.319,22	$2.290,45	$15,51	$28,77	$2.334,73	$9.203,70
357	$2.315,32	$2.292,31	$11,63	$23,01	$2.326,95	$6.899,76
358	$2.311,43	$2.294,18	$7,75	$17,25	$2.319,18	$4.597,83
359	$2.307,54	$2.296,05	$3,87	$11,49	$2.311,41	$2.297,91
360	$2.303,66	$2.297,91	$0,00	$5,74	$2.303,66	$0,00

Exemplary Ammortization Table of a $1,000,000 Loan Pool with a CPR of 2%

Loans Distribution by MSA

MSA Code	MSA Name	Number of Loans
12060	Atlanta, GA	1176
14460	Boston, MA	1189
16740	Charlotte, NC	473
16980	Chicago, IL	1665
17420	Cleveland, OH	15
19100	Dallas, TX	1708
19740	Denver, CO	926
19820	Detroit, MI	696
29820	Las Vegas, NV	880
31080/31100	Los Angeles, CA	4981
33100	Miami, FL	1125
33460	Minneapolis, MN	814
35620	New York City, NY	5965
38060	Phoenix, AZ	1946
38860	Portland, OR	90
41740	San Diego, CA	1428
41860	San Francisco, CA	2015
42660	Seattle, WA	1158
45300	Tempa, FL	626
47900	Washington, D.C.	2285

Loans Distribution by Region

Region	Number of Loans
North-Central	3190
North-East	9439
South-East	3400
South-West	3654
West	11479