Modeling Defaults in Residential Mortgage Backed Securities: An Intensity Based Approach

by

Toma Donchev

August 2009

Supervision:

Prof. dr. Aad van der Vaart

Dr. Federico Camia

Jurgen Peters

Barbara Bakker

Submitted to the Vrije Universiteit Amsterdam as a requirement for the successful graduation in the degree:

Master of Science in *Business Mathematics and Informatics* – specialization *Financial Risk Management*

Modeling Defaults in Residential Mortgage Backed Securities: An Intensity Based Approach

Abstract

In May 2008 the outstanding issuance of European Asset Backed Securities was more than €1150 billion and Residential Mortgage Backed Securities (RMBS) accounted for 77% of this amount. Naturally, given these figures, managing default risk of the collateral pool becomes of crucial importance to financial institutions and investors. In this paper we present an intensity based approach for modeling residential mortgage defaults. More specifically, we will fit a Cox proportional hazard rate model to describe the probability of default (PD) for residential mortgages and the uncertainty around the expected PD. Once we are able to model mortgage defaults we will turn our attention to modeling the distribution of loss given default (LGD) and determine the effects of defaults to the RMBS tranches. We hope that the results of this research will improve NIBC's existing methods for managing the credit risk originating from the collateral pool of European RMBS transactions.

Contents:

I. Introduction

1.1 An overview of Residential Mortgage Backed Securities

Residential Mortgage Backed Securities (RMBS) are financial securities backed by a pool of residential mortgages. The process of creating RMBS (as well as all other types of structured credits) is called securitization. In this section we present a short description of the RMBS transactions and the risks associated with them.

Residential Mortgage Backed Securities are structured credits that can be characterized by the following: the originator (usually a bank) has a pool of residential mortgages on its balance sheet. The originator sells those to a so called *Special Purpose Vehicle* (SPV), a company created solely for the purpose of securitization. The SPV raises funds to purchase these mortgages by issuing notes to investors. In this way the investors only bear the risk arising from the pool of mortgages (collateral pool) and are generally independent from the credit risk of the respective (former) owner of those assets (e.g. originating bank).

The assets (in this case residential mortgages) of the collateral pool generate interest and principal payments. These payments as well as potential losses, that may occur in case the underlying borrowers do not serve their obligations, are distributed to the investors according to the structure of the securitization. In this way the credit risk of the collateral pool is transferred to the investors. The notes are divided into several classes with different seniority, varying from AAA to Equity. In general the notes with the lowest rating are the first to absorb losses in

the underlying pool of assets. Naturally, the notes with lowest rating have the highest risk and accordingly generate the highest return. Respectively, the most senior notes are the least risky ones and produce the lowest return. The set of rules, which distributes cash flows (and losses) from the collateral to the notes, is called the *waterfall* of the structured credit. Each RMBS deal has its specific waterfall. Therefore, investors in RMBS have to focus on both the underlying risk of the securitized portfolio (collateral pool) and the rules that determine which consequences investors have to face in case certain events occur. The fact that different notes have different risk profiles, though they all reference the same underlying portfolio, is based on the respective special transaction structure. This enables investors to satisfy their individual risk appetites and needs. Figure 1.1 depicts the general structure of a typical RMBS transaction.
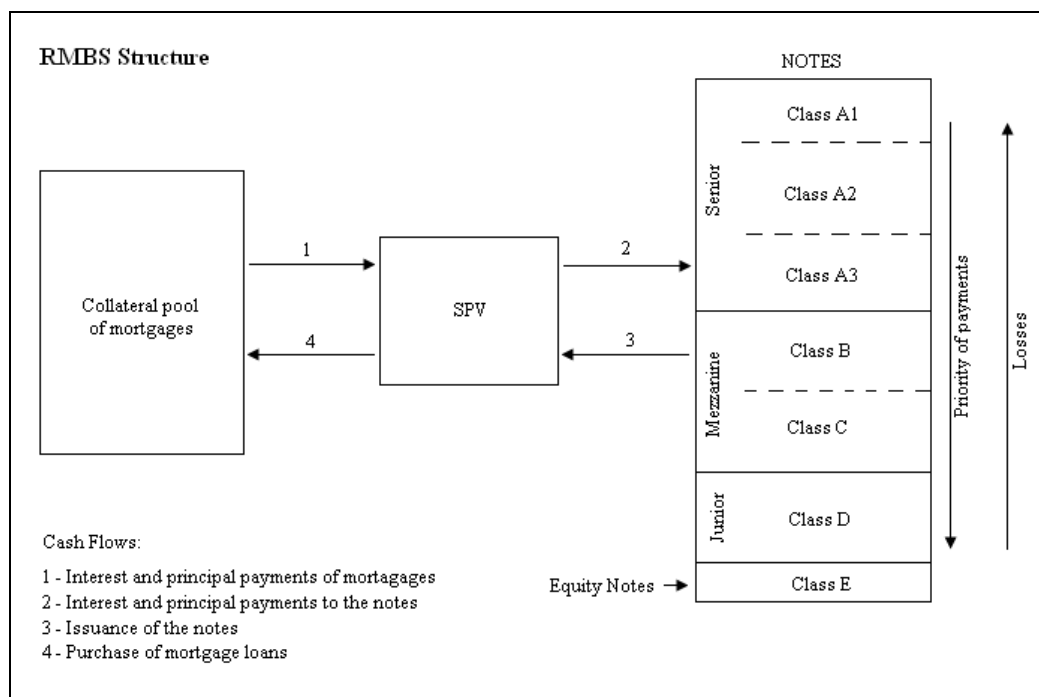
Figure 1.1 RMBS General Structure

Because the waterfall of each RMBS deal is unique and has been determined at origination, the effect of a given (expected) loss from the pool to the notes is specific for each RMBS but deterministic. NIBC's Trading Department has its own model to determine this effect deterministically. On the other hand, from a risk management point of view it is important to have a model that stochastically describes the uncertainty of the losses originating from the pool i.e. the uncertainty around the expected PD. In this paper we will present a stochastic approach for modeling the credit risks associated with the collateral pool of residential mortgages.

1.2   Mortgage Termination – Default or Prepayment

A great deal of research exists today on modeling mortgage termination. A general consensus exists in the literature – a mortgage is terminated if it is either prepaid or the borrower has defaulted from his payment obligations (*Deng* [3] and *Deng, Quigley & Van Order* [4]).

The goal of this paper, however, is to model the uncertainty around the expected loss associated with RMBS securities. In structured credits the proceeds from prepaid (and also paid on their legal maturity) mortgages are used by the SPV to either replenish the mortgage pool (purchase new mortgages), or to repay some of the outstanding notes. In the first case there is practically no effect of prepaid mortgages to the cash flows of the notes. In the second case this effect is determined by the specific RMBS contract - repayment of notes could be for example proportional to the notes tranches (in this case again there is practically

no impact to the credit enhancements of the notes) or it could only affect the most senior notes. To summarize - in RMBS transactions, prepayments either have no impact on the cash flows to the notes or this impact is deterministic (determined by the waterfall). Therefore in the scope of this paper, the credit risk associated with mortgage portfolios is essentially the risk that borrowers will default and fail to meet interest rate payments on the outstanding balance plus the risk that given default, the collateral value of the defaulted mortgage is less than the outstanding balance plus unpaid interest.

1.3   Structural vs. Intensity Based Approach

The credit risk modeling literature has been essentially developed in two ways – the structural approach and the reduced-form approach. The structural approach is also sometimes called option-based approach. The ancestor of all structural models is the *Merton Model* [5]. The main idea is to use the evolution of firms' (borrowers') structural variables, such as asset (house) and debt values, to determine the time to default or to prepay. Default is viewed as a put option; the borrower sells his house back to the lender in exchange for eliminating the mortgage obligation. Whereas, prepayment is viewed as a call option; the borrower exchanges the unpaid balance on the debt instrument for a release from further obligation. In the structural approach it is assumed that there are no transactional or reputation costs for default or prepayment and that, borrowers are well-informed and make the rational choice to exercise either the call or the put option when they can increase their wealth. These assumptions may look appropriate when dealing with commercial borrowers but are not realistic when considering

residential mortgages. The behavior of private individuals, whose purpose is to finance their property with the loan, is not always rational in the sense of the economic theory.

Another shortcoming of structural models arises when considering the legal aspects of mortgage contracts. The majority of structural models were developed in attempt to describe the credit risk of the mortgage market in the United States. While in the US, borrower's obligations to the originator of the loan are terminated in the case of default (the bank only has rights on the property, no matter if its actual market value is less than the value of the mortgage contract), this is not the case in Europe. In most European residential mortgage contracts, if a borrower defaults he loses his property and if the market value of the property does not cover the present value of the outstanding interest payments the borrower is also obliged to cover this difference. In this case, the approach of modeling default behavior as a put option on the house value is quite unrealistic.

An extensive literature exists employing the structural approach in the valuation of mortgages (see, for instance, *Titman & Torous* [6]*, Kau et al* [7] or *Kau and Keenan* [8]). While the option based viewpoint has yielded considerable insights into the workings of idealized mortgages, it has proven difficult to employ such models for the purpose of empirical estimation.

The reduced-form models are also called *intensity* or *hazard rate* models. Compared to structural reasoning, the reduced-form point of view is a good deal less economical: default or prepayment is no longer internally determined, but rather, externally imposed on the model according to some random process. In intensity based models, the default time is modeled as a first jump time of an

exogenously given jump process. In the literature several explanatory variables for a default of a mortgage contract have been identified. *Smith, Sanchez and Lawrence* [9] and *Deng* [3] select mortgage specific and economic characteristics for predicting defaults and for calculating the probability of incurring a loss on a defaulted loan. *Santos Silva and Murteira* [10] use borrower's characteristics, such as the Debt-To-Income ratio (DTI), which is usually only observable by the issue of the mortgage. In their model, *Follian, Huang, and Ondrich* [11] include duration, location, demographic and economic variables as covariates to explain default.

Combinations of the structural and reduced form models also exist. To model time to default, *Deng* [3] and *Deng and Quigley* [12] propose combining the financial value of the put option in the structural approach, with non-option related variables, such as unemployment or divorce rates.

In this paper we consider residential mortgage default as an event which is triggered by mortgage specific, macro-economic and by some personal "non-financial" reasons, more than by a rational economic decision (see also *Deng and Quigley* [12] and *De Giorgi* [13]). One common (macro-economic) cause for default is unemployment; another is divorce. In the case of unemployment the income of the borrower can dramatically decrease and the consequence will be the inability to pay the interest on the outstanding balance. Therefore, considering an RMBS collateral pool, we try to model the distribution of the expected number of defaults according to the economic environment, i.e. to economic factors such as unemployment and interest rates, or to social and demographic developments, such as the increase of the number of divorces.

We propose an intensity based approach for modeling the time to default, which we take to be the first-jump-time of an inhomogeneous Poisson process with stochastic intensity, also called a doubly stochastic Poisson or Cox Process. The main idea consists in conditioning on a set of explaining variables (e.g. loan-to-value (LTV) ratio or DTI), which affect borrowers' *credit quality* and behavior, and to consider borrower defaults as independent given the set of information about the common economic environment. The intensity process is directly related to the underlying explanatory variables, as in the *proportional hazard rate model* (PHR) of *Cox and Oakes* [14]. The specific characteristic of the model are presented in the next chapters.

1.4   Loss Given Default Models

As already mentioned above – a mortgage contract will cause losses if given default, the collateral value of the defaulted mortgage is less than the outstanding balance plus unpaid interest. We therefore also need a way to model the *loss given default* (LGD) of the residential mortgages in the collateral pool.

In the existing credit risk literature, initial approaches for LGD estimation were deterministic in nature. Nevertheless, nowadays it has become widely accepted to treat LGD as a loss severity distribution rather than to regard each estimate as deterministic, since a number of factors play a role in the ultimate recovery, and to estimate these deterministically is a difficult task.

There are two main approaches for modeling loss severity. The first one is deterministic – it simply assumes that all recoveries are fixed values that are known in advance. The argument for this simplification is the facts that the

uncertainty of the recovery rates does not contribute significantly to the risk of losses, when compared with the default rate volatility. In other words, the default rate estimate dominates the LGD estimate, when estimating the expected loss of an exposure.

The second method models the recovery rates as a random variable between 0% and 100%. The LGD of a mortgage is then given as 1 minus the recovery rate. Most often in the literature, a U-shaped *beta* distribution is used to model the recovery values. This distribution is very useful because it can be bound between two points and can assume a wide range of shapes. Many popular commercially available portfolio management applications use a beta distribution to model the recovery value in the event of default. In this paper we adopt this methodology to model the LGD of RMBS's collateral pool of mortgages. The exact estimation procedure and the calibrating of the beta distribution are presented in the next chapter.

The rest of the paper is organized as follows: section II gives a short mathematical background of random times, hazard rates and jump processes; in section III the model for probability of default is introduced; section IV describes the estimation methodology, the availability of data and the loss-given-default model; in section V we present the results of our simulations and section VI concludes the study. Technical results are reported in the Appendix.

II. Mathematical tools

Let us first start by presenting some mathematical tools for the analysis of reduced-form models. In particular we will focus on random times and hazard rates. We start with random times with deterministic hazard rates and after that we consider situations where the only observable quantity is the default time itself. This forms the basis for an analysis of a more realistic situation where additional information, generated for instance by economic explanatory variables, is available, so the hazard rate will typically be stochastic. We give a description of the doubly stochastic random times. Doubly stochastic random times are the simplest example of random times with stochastic hazard rates and are thus frequently used in dynamic credit risk models. We assume that the reader is familiar with the basic notions from the theory of stochastic processes, such as filtration, stopping times or basic martingale theory.

2.1 Random Times and Hazard Rates

Let us consider a probability space $(\Omega, \mathbf{F}, \mathbf{P})$ and a *random time* τ defined on this space i.e. $\tau : \Omega \to (0, \infty)$ is a positive, $\mathbf{F}$-measurable random variable that is interpreted as the time to default of a mortgagor. We denote by $F(t) = P(\tau \leq t)$ the cumulative distribution function of τ and by $\overline{F}(t) = 1 - F(t) = P(\tau > t)$ the survival function of τ. We assume that $F(0) = P(\tau = 0) = 0$ and that $\overline{F}(t) > 0$ for all $t < \infty$. We can now define the jump or default indicator process (Y_t) associated with τ by

$Y_t = I_{\{\tau \le t\}}$ for $t \ge 0$. Note that (Y_t) is a right continuous process which jumps from 0 to 1 at the default time τ and that $1 - Y_t = I_{\{\tau > t\}}$.

A *filtration* $(\mathbf{F}_t)$ on $(\Omega, \mathbf{F})$ is an increasing family $\{\mathbf{F}_t : t \ge 0\}$ of sub-σ-algebras of $\mathbf{F}$: $\mathbf{F}_t \subset \mathbf{F}_s \subset \mathbf{F}$ for $0 \le t \le s < \infty$. For a generic filtration $(\mathbf{F}_t)$ we set $\mathbf{F}_\infty = \sigma(\bigcup_{t \ge 0} \mathbf{F}_t)$. In practice filtrations are typically used to model the flow of information. The filtration $\mathbf{F}_t$ represents the state of knowledge of an observer at time t and $A \in \mathbf{F}_t$ means that at time t the observer is able to determine if an event A occurred.

In the following we assume that the only observable quantity is the random time τ and equivalently the associated jump process (Y_t). Let $(\mathbf{H}_t)$ be given by

$$\mathbf{H}_t = \sigma(\{Y_u : u \le t\}) \tag{2.1}$$

$(\mathbf{H}_t)$ is the filtration generated by default indicator process i.e. the history of the default information up to and including time t. By definition, τ is an $(\mathbf{H}_t)$-stopping time as $\{\tau \le t\} = \{Y_t = 1\} \in \mathbf{H}_t$ for all $t \ge 0$.

Definition 2.1 (hazard rates and cumulative hazard function)

The function $\Gamma(t) := -\ln(\overline{F}(t))$ is called the cumulative hazard function of the random time τ. If F is absolutely continuous with density f, the function $\gamma(t) := f(t)/(1 - F(t)) = f(t)/\overline{F}(t)$ is called the hazard rate of τ.

By definition we have $F(t) = 1 - e^{-\Gamma(t)}$ and $\Gamma'(t) = f(t)/\overline{F}(t) = \gamma(t)$, so $\Gamma(t) = \int_0^t \gamma(s)\, ds$. When we consider a very small interval of time, the hazard rate $\gamma(t)$ can be interpreted as the instantaneous chance of default at time t, given survival up to time t. For $h > 0$ we have:

$$P(\tau \le t + h \mid \tau > t) = \frac{F(t+h) - F(t)}{1 - F(t)} \tag{2.2}$$

and therefore

$$\lim_{h \to 0} \frac{1}{h} P(\tau \le t + h \mid \tau > t) = \frac{1}{\overline{F}(t)} \lim_{h \to 0} \frac{F(t+h) - F(t)}{h} = \gamma(t) \tag{2.3}$$

The hazard rate $\gamma(t)$ can be interpreted as the expected number of failures (defaults) in a unit of time. Since integration is practically summation – the cumulative hazard function $\Gamma(t) = \int_0^t \gamma(s)\, ds$ can be understood as the expected number of failures in the period of time between 0 to *t.*

There are several advantages in learning to think in terms of hazard rates, rather than the traditional density functions and cumulative distribution functions. Hazard functions give a more intuitive way to interpret and understand the process that generates failures. This is why in survival analysis regression models are more easily grasped by explaining how different variables (covariates) affect the hazard rate.

III. Model for Probability of Default

As already stated, in this paper we will try to describe the probability of default of residential mortgages via an intensity based model. Moreover, our goal is to quantify the dependence and sensitivity of the PD on some explanatory factors. These factors can be mortgage specific (LTV and/or DTI ratios) or external (unemployment and/or interest rates). Most of the intensity-based models, including ours, maintain a *doubly stochastic* character, which means that not only it is uncertain whether an obligor will default at a particular time, but that the intensity (hazard rate) by which this event occurs is also uncertain beforehand. Therefore we try to model the default time of a residential mortgage as a random time with a stochastic hazard rate. This leads to the use of the so-called *doubly stochastic Poisson Process* to model the probability of default.

3.1 Duration and Time to Default

Consider the following setting. Let $P = \{(s_i, B_i, V_i), i = 1,...,n\}$ be a portfolio of n residential mortgages. For mortgage i, s_i denotes the time of issue (calendar time), $B_i = (B_{i,t})_{t \geq s_i}$ is a process giving the outstanding balance at time t and $V_i = (V_{i,t})_{t \geq s_i}$ is a stochastic process giving the house value at time t. We suppose that the mortgage portfolio is totally characterized by P.
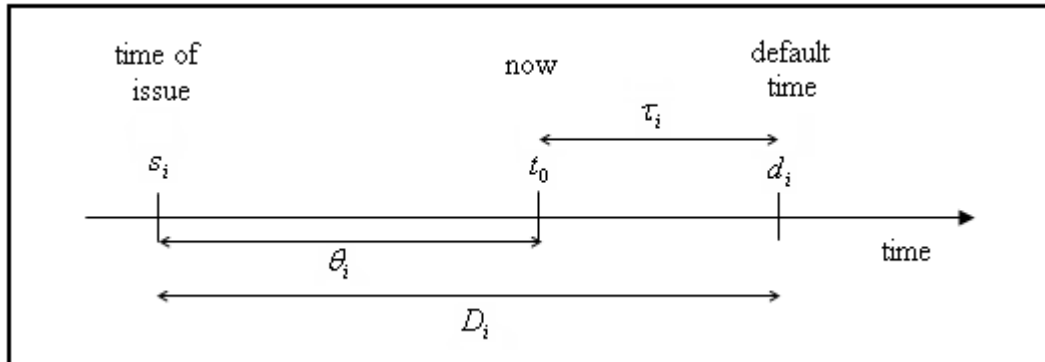
Figure 3.1 Duration and time to default.

Now let $D_i : \Omega \to (0, \infty)$ be a positive random variable giving the duration or lifetime of a mortgage contract i and let $\tau_i : \Omega \to (0, \infty)$ be also a positive random variable giving the time to default of a mortgage i i.e. τ_i is the period of time from now (t_0) till the obligor i defaults. We assume that for all mortgage contracts in P we have $\mathbf{P}(D_i = 0) = 0$. Moreover, $\mathbf{P}(D_i > d) > 0$, $\forall d > 0$ and also $\mathbf{P}(\tau_i = 0) = 0$ and $\mathbf{P}(\tau_i > t) > 0$, $\forall t > 0$. We have $D_i = \infty$ (or equivalently $\tau_i = \infty$) if mortgage i does not default. Also let d_i denote the default time (calendar time) of obligor i and θ_i be the period of time that the mortgage has been outstanding (period of time from issue till now – current lifetime of the mortgage). See figure 3.1.

Since s_i is known in advance for any given time t_0, we can calculate θ_i and since θ_i is known and deterministic, the distribution (hazard rate) of time to default τ_i is completely determined by the distribution (hazard rate) of the duration D_i. Mathematically we have:

$$P(\tau_i \leq t) = P(D_i \leq \theta_i + t \mid D_i > \theta_i) \tag{3.1}$$

In other words – the probability that mortgage i will default in a certain interval of time t is equal to the probability that the lifetime of the mortgage i is less than or equal to the current lifetime of the mortgage plus t, given that the mortgage is still outstanding at time t_0 (has survived till t_0). Moreover we have:

$$P(\tau_i \le t) = P(D_i \le \theta_i + t \mid D_i > \theta_i) = \frac{P(D_i \le \theta_i + t) - P(D_i \le \theta_i)}{P(D_i > \theta_i)} \qquad (3.2)$$

and if we know the hazard rate or the cumulative hazard function for the distribution of D_i then all the values on the right hand-side of (3.2) are known.

We will now use Cox Proportional Hazard Rate Model to model the hazard rate of mortgage duration.

3.2   The Cox Proportional Hazards Model

In this chapter we present a way to model (and later estimate) the hazard rate of mortgage duration. We borrow a model typically used in medical science in the field of survival analysis. The Cox Model is a well-recognized statistical technique for exploring the relationship between the survival of a patient and several explanatory variables (also called covariates).

In our case we assume that mortgage defaults are triggered by some mortgage specific and/or by some external (environment specific) factors. We suppose that we can find a set of predictors for the default event of obligor i. Mathematically we have a multi-dimensional stochastic process $\mathbf{X}_i = (X_1, ..., X_p)$, such that each

component $X_{i,q}$ $(q=1,...,p)$ represents an explaining factor for the event of default of obligor i, as for example the regional unemployment rate.

Let $\lambda(t\,|\,\mathbf{X}_i)$ be the hazard rate of mortgage duration, given a particular realization of the default factors $\mathbf{X}_i = (X_1,...,X_p)$. Note that $\lambda(t\,|\,\mathbf{X}_i)$ simply states that $\lambda(t)$ is a function of $\mathbf{X}_i = (X_1,...,X_p)$. Cox Proportional Hazard Model assumes that the relationship between $\lambda(t\,|\,\mathbf{X}_i)$ and the explanatory factors $\mathbf{X}_i$ is given by:

$$\lambda(t\,|\,\mathbf{X}_i) = h(t) * \exp(\beta'\mathbf{X}_i) \tag{3.3}$$

where $h(t)$ is the baseline hazard (effect of (elapsed) time t on mortgage duration) and β is a vector of coefficients giving the sensitivity of the hazard rate to changes in the explanatory variables. Note that the baseline hazard $h(t)$ corresponds to the instantaneous probability of default, given survival (no default) up to time t when all the covariates are zero. The model also states that the baseline function is the same for all mortgages in consideration i.e. the default rates of mortgage i and mortgage j differ only in the realizations of the covariates $\mathbf{X}_i$ and $\mathbf{X}_j$. This fact plays a crucial role in the estimation procedure, as we will see later. From (3.3) and the definition of the hazard rate, it follows that the cumulative hazard function of mortgage duration is equal to:

$$\Lambda(t\,|\,\mathbf{X}_i) = \int_0^t h(s) * \exp(\beta'\mathbf{X}_i)\,ds \;=\; \exp(\beta'\mathbf{X}_i)\int_0^t h(s)\,ds = \exp(\beta'\mathbf{X}_i)H(t) \tag{3.4}$$

where $H(t) = \int_0^t h(s)\, ds$ is the baseline cumulative hazard function.

The model explains the following behavior of the default intensity process: suppose that at the beginning of the mortgage agreement an expected intensity (default rate) λ_0 can be associated to obligor i. If the obligor's behavior is not affected by any predictors $X_1,...,X_p$, then we expect no contribution of $\mathbf{X}_i = (X_1,...,X_p)$ to the intensity process, meaning that $\beta_i = 0$, for all $i = 1,...,p$. Moreover, if the elapsed time does not contribute to the default intensity, then $\lambda(t) = h(t) \equiv \lambda_0$ is constant, which would imply a homogenous Poisson process.

However, in practice we observe that obligor's behavior changes during the life of the mortgage, meaning that the probability of incurring a default increases or decreases. Some factors $X_1,...,X_p$ affect the ability of obligor i to pay the interest rate on a mortgage, changing stochastically the default intensity. Equation (3.2) suggests that predictors $X_1,...,X_p$ and time t affect the realizations of $\lambda(t)$ in a multiplicative way.

Another thing worth mentioning is default correlation, which is certainly lower for residential mortgages compared to commercial mortgages. Dependence between residential mortgage defaults can be explained, to a large extend, only by the macroeconomic environment (e.g. unemployment rate and/or interest rate). This allows us to assume conditional independence of residential defaults. We assume that obligors who default up to time t are conditionally independent, given the history of the predictors up to time t. This assumption seems reasonable for

the kind of portfolio we are considering in this paper – a portfolio of private individuals (for company defaults and commercial mortgages this assumption would not be realistic). In fact this conditional independence implies that, given a scenario through the predictors, obligor defaults occur independently, meaning that the dependence structure is fully described by the evolution of the common (macroeconomic) covariates.

IV.   Available Data and Model Estimation

4.1   Residential Mortgage Historical Data

NIBC Bank N.V. has maintained a significant database of Dutch residential mortgages. The database contains approximately 92 thousand records and was recorded between 01/01/2002 and 6/1/2008 so all still existing contracts have *"end_date" - 6*/1/2008. See table 4.1 for a sample of the database.

Original LTV	*Original DTI*	*Loan start date*	*Loan end date*	*Right censored data (0=defaulted)*
0.390438728	0.218330602	7/1/2000	6/1/2008	1
1.154709643	0.141826843	3/1/2005	6/1/2008	1
1.14553197	0.269474451	3/1/2005	6/1/2008	1
1.158583728	0.37568185	3/1/2005	6/1/2008	1
0.808333333	0.126808706	3/1/2008	6/1/2008	1
1.285714286	0.1235684	3/1/2005	10/1/2006	0
0.966666667	0.286158458	3/1/2005	6/1/2008	1
1	0.170747839	3/1/2005	6/1/2008	1

1.343134328	0.367103444	1/1/2006	6/1/2008	1
0.642487047	0.121560317	3/1/2007	6/1/2008	1
0.898305085	0.224445451	3/1/2005	6/1/2008	1
0.423536	0.235642	3/1/2005	10/1/2007	0
1.229367273	0.303740859	3/1/2005	6/1/2008	1
1.167959184	0.248989162	3/1/2005	6/1/2008	1

Table 4.1 Sample from the Mortgage Database

Each row represents one mortgage contract. The last column shows the default status of a mortgage loan i – it is 0 if obligor i has defaulted (and contract seized to exist in the database at its *end_date*); and it is 1 if mortgage i is either still existing or it was terminated due to prepayment or repayment (and was removed from the database on this corresponding *end_date).* A mortgage contract is considered to have defaulted when it has been in arrears for more than 3 months i.e. the obligor has made no interest or principal payments on his mortgage obligation for more than 3 months. As default is an extremely rear event (especially in the Netherlands), almost all of the observations are censored (last column is 1). In fact for the 6 years in consideration there were only 1558 defaults out of 92 thousand mortgage loans.

The duration (lifetime) of mortgage contract i is obtained by taking the difference between *end_date i* and *start_date i* , expressed in months.

4.2 Default Predictors

As we already stated, the realization of the covariates (default predicting factors) has significant impact on the realization of the hazard rate of mortgage duration. To construct our model we chose 4 default factors, namely:

$X_{1,i}$	Original Loan-To-Value (LTV) Ratio of loan i
$X_{2,i}$	Original Debt-To-Income Ratio (DTI) of loan i
$X_{3,i}$	Quarterly unemployment rate (at contract's *end date*)
$X_{4,i}$	3-month Euribor interest rate (at contract's *end date*)

Table 4.2 Default Predictors

LTV and DTI are a common choice for factors explaining mortgage defaults. LTV stands for *loan-to-value ratio* and gives the ratio of the size of the mortgage loan to the value of the real estate property – or simply – loan value over house value. The original LTV is the loan-to-value ratio of borrower i at origination of the mortgage contract. As we will show later LTV has very small and statistically insignificant impact on the hazard rate of time to default. The DTI ratio on the other hand has a significant explanatory power in our model. It stands for *debt-to-income ratio* (sometimes also called PTI (*payment-to-income*) ratio) and expresses the ratio of monthly payments due on the mortgage loan to the reported income of borrower i i.e. it directly relates the payment weight to the ability of payment for obligor i. As one would expect, the DTI ratio has a significant importance in explaining the hazard rate of time to default.

We also choose Quarterly Unemployment Rate and Euribor Interest Rate (with the appropriate lag – this will be explained later), because they are macroeconomic variables that should have an impact on obligors' ability to pay the interest on their mortgage obligations. A rise in unemployment will mean that more people lose their primary source of income which will affect their ability to pay interest on their loans. Same is true for interest rates – for most Dutch residential mortgages (and most of the mortgages in our database) the interest payments due to the obligor are determined by a base interest rate (Euribor) plus a margin. This means that a large increase in Euribor rates will increase the payment weight of mortgagors and consequently will make some obligors incapable of paying these payments.

We have to point out here that there are two definitions of unemployment rate in the Netherlands. One is expressed as a percentage of total population and one – as a percentage of the labor force (that is population between 16 and 65 years of age). As expected they are almost perfectly correlated with each other and it makes practically no difference which one we use in our model. The only difference is the β-coefficient for unemployment in the Cox regression model (3.3). We are going to use the one that is more frequently used in the media and namely the one that is expressed as a percentage of the labor force.

For contracts that have been terminated during the period of our study, we assign unemployment rate and the interest rate at the month of termination. Looking at the data we see that we have a large amount of mortgages that are still outstanding (approximately two thirds of the records) and their actual covariates ($X_{3,i}$ and $X_{4,i}$) are not observed. We have no actual *end_date* for mortgage

contracts that still exist. To overcome this problem with missing values, we simply assign a 0 for $X_{3,i}$ and $X_{4,i}$ for contracts that are still existing. In the next section we will explain why we make this choice and its impact on the estimation of the β coefficients.

4.3   PD Model Estimation

In this section we explain the mathematics behind the estimation of the Cox Model. A non-parametric method for estimating the β coefficients was developed by *Cox* [15] himself and is called *partial likelihood* estimation. The estimation is non-parametric, meaning that the baseline hazard can be left unspecified. This means that we do not have to assume a certain shape for the baseline function. In this way the estimation is not biased by the choice of a baseline hazard.

Suppose we have a data set with n observations and k distinct failure (event or default) times. We first sort the ordered failure times such that $t_1 < t_2 < ... < t_k$, where t_i denotes the failure time of the i-th mortgage. Note that $t_1 < t_2 < ... < t_k$ are the actual times when default happened i.e. only uncensored cases (last column in our database = 0). We now want to express the event times as a function of the covariate matrix $\mathbf{X}$.

The partial likelihood function is derived by taking the product of the conditional probability of a failure at time t_i, given the number of mortgages at

risk at time t_i. In other words, given that a default has occurred, what is the probability that it occurred to the i-th mortgage from a risk set of size N?

Let $R(t_i)$ denote the number of mortgages that are at risk of failing (defaulting) at time t_i i.e. $R(t_i)$ is the relevant risk set. Then the probability that the j-th mortgage will default at time t_i is given by:

$$P(t_j = t_i \mid R(t_i)) = \frac{\lambda(t \mid \mathbf{X}_i)}{\sum\limits_{j \in R(t_i)} \lambda(t \mid \mathbf{X}_j)} = \frac{h(t)*\exp(\beta'\mathbf{X}_i)}{\sum\limits_{j \in R(t_i)} h(t)*\exp(\beta'\mathbf{X}_j)} = \frac{\exp(\beta'\mathbf{X}_i)}{\sum\limits_{j \in R(t_i)} \exp(\beta'\mathbf{X}_j)} \qquad (4.1)$$

since the baseline hazard $h(t)$ is the same for all mortgages. The denominator in the above expression is the summation over all mortgages that are at risk at time t_i. Taking the product of these conditional probabilities yields the partial likelihood function:

$$L_p = \prod_{i=1}^{k} \left[\frac{\exp(\beta'\mathbf{X}_i)}{\sum\limits_{j \in R(t_i)} \exp(\beta'\mathbf{X}_j)} \right] \qquad (4.2)$$

with corresponding log-likelihood function:

$$\log L_p = \sum_{i=1}^{k} \left[\beta'\mathbf{X}_i - \log\left(\sum_{j \in R(t_i)} \exp(\beta'\mathbf{X}_j) \right) \right] \qquad (4.3)$$

By maximizing the log-likelihood function (4.3), estimates of β are obtained.

Note that the product in (4.2) and the sum in (4.3) are over all mortgages i that have actually defaulted i.e. $i = 1,....k$ and $k = 1558$ in our case. Those 1558 are exactly the contracts that are not censored, therefore for all $i = 1,....k$, we have $X_{3,i} \neq 0$ and $X_{4,i} \neq 0$ (as mentioned before, for all mortgages that still exist - no default event has occurred - we set $X_{3,i} = 0$ and $X_{4,i} = 0$).

Of course the risks sets $R(t_i)$, $i = 1,....k$ contain all mortgages that are at risk of defaulting at time t_i which includes censored cases and consequently mortgages that are still existing (for which we have no actual observed unemployment and interest rate values). From (4.1) we see that censored cases contribute information only relevant to the risk set (denominator of (4.1) and (4.2)). Therefore by setting $X_{3,j} = 0$ and $X_{4,j} = 0$ for all those still existing mortgages we actually set $\exp(\beta' \mathbf{X}_j) \approx 1$ (because $\beta' \mathbf{X}_j \approx 0$) and the denominator in (4.1) and (4.2) is simply increased by 1 for each mortgage contract that still exists. In this way we try to minimize any bias coming from the fact that we are unable to observe the unemployment and interest rate values of the month of termination of all those still existing loans. On the other hand we can not simply remove them from the estimation because we do not want to lose any information about the occurrence of defaults (and the fact that default is an extremely rear event). Mathematically by setting $X_{3,i} = 0$ and $X_{4,i} = 0$, we remove any effects of those unobserved variables to the weight of the risk set – the denominator of (4.1) and (4.2), and we let the hazard rate for these contracts be determined only by their baseline hazard

and the other two constant (and known for all contracts) covariates X_1 (loan-to-value ratio) and X_2 (debt-to-income ratio).

Once we have estimated the β coefficients we also need an estimate of the baseline hazard $h(t)$ to finally obtain an estimate of the hazard rate of failure $\lambda(t)$. In the literature there are a number of approaches that have been adopted to estimate the baseline hazard $h(t)$. The simplest and most frequently used approach was proposed by *Breslow* [16]. He derived a maximum likelihood estimator of the baseline cumulative hazard function $H(t) = \int_0^t h(s)\,ds$, after assuming that the failure time distribution has a hazard rate which is constant between each pair of successive observed failure times - a reasonable assumption in our case. The estimate of $h(t)$ in the interval $[t_{i-1}, t_i]$ between two successive failure times t_{i-1} and t_i is given by:

$$\tilde{h}_i = \frac{d_i}{\delta_i \sum\limits_{j \in R(t_i)} \exp(\beta' \mathbf{X}_j)} \tag{4.4}$$

where $\delta_i = t_i - t_{i-1}$ is the length of the time interval and d_i is the number of defaults that occur in time t_i (note that usually $d_i = 1$, but this estimation also allows for multiple failures at the same time t_i - these failures are also called *ties*). Equation (4.4) can be interpreted as the ratio between the number of events and the weighted number of *'person-time'* units at risk, where the weight of each individual j in the risk set $R(t_i)$ is $\exp(\beta' \mathbf{X}_j)$. A rough estimate of $H(t_i) - H(t_{i-1})$

Breslow's estimator of the cumulative baseline hazard function at time t:

$$\tilde{H}(t) = \sum_{t_{(i)} \leq t} \frac{d_i}{\sum_{j \in R(t_i)} \exp(\beta' \mathbf{X}_j)} \qquad (4.5)$$

4.4   Modeling Loss Given Default

Residential mortgage loans are always backed by some kind of real estate collateral. If an obligor defaults on his payment obligations then the lender gets hold of the collateral. The recovery value – i.e. proceeds from selling this collateral expressed as a percentage of loan's outstanding balance, are used to cover losses arising from defaults of obligors.

As we already mentioned we want to model the recovery rates as a random variable between 0% and 100%. The _loss-given-default_ (LGD) of a mortgage is then given as 1 minus the recovery rate. Most often in literature, a U-shaped _beta_ distribution is used to model the recovery values. The beta distribution is very useful for modeling recovery rates because it produces values between 0 and 1 and can have a large variety of shapes (see fig.4.1).

The probability density function of the Beta distribution is given by:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad (4.6)$$

$$= \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ is the gamma function.
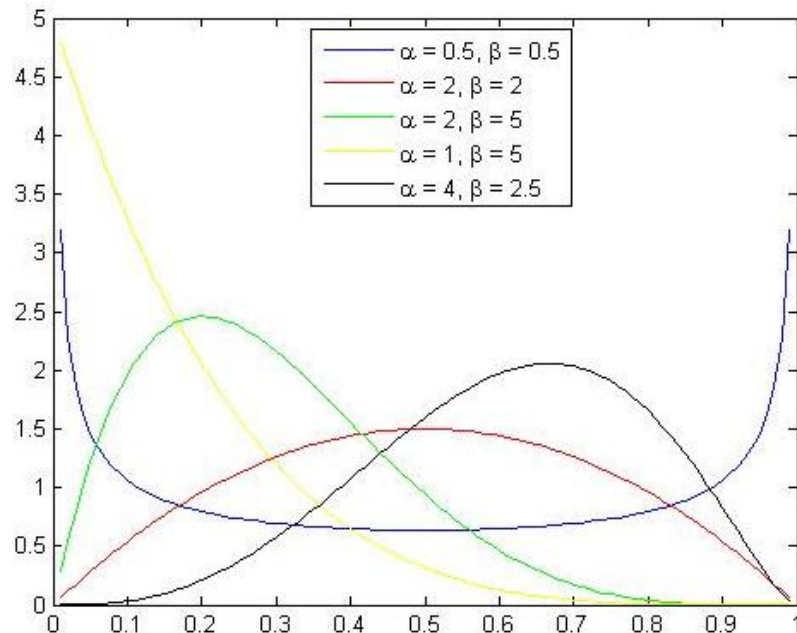
Figure 4.1 The PDF of the beta distribution for different values of the parameters α and β:

As we see from fig. 4.1 the shape of the Beta distribution is determined by the parameters α and β. These parameters are usually estimated in the following way:

$$\alpha = \mu \cdot \left(\frac{\mu \cdot (1-\mu)}{\sigma^2} - 1 \right) \text{ and } \beta = (1-\mu) \cdot \left(\frac{\mu \cdot (1-\mu)}{\sigma^2} - 1 \right) \quad (4.7)$$

where μ and σ are the mean and standard deviations of the recovery rates. NIBC Bank N.V. has a data set of residential mortgages from which we obtain the following characteristics of recovery rates:

Number of Losses	Average RR	Std. Dev. RR	Average LGD
860	89.59%	19.19%	10.41%

and the α and β parameters become: $\alpha = \mu \cdot \left(\frac{\mu \cdot (1-\mu)}{\sigma^2} - 1 \right) = 1.7185$ and

$$\beta = (1-\mu) \cdot \left(\frac{\mu \cdot (1-\mu)}{\sigma^2} - 1 \right) = 0.4537.$$

Once we have calibrated the correct Beta distribution we can use it to simulate random recovery rates and combining those with the distribution of the expected probability of default, obtained from our PD model, we can determine the distribution of the expected loss due to default of a single mortgage contract or of a portfolio of mortgage loans.

V. Results

5.1 Cox PH model estimation and regression results

In the previous chapter we proposed using 4 distinct default predictors (*Original LTV, Original DTI, Unemployment rate* and *Interest rate*) for building the most suitable proportional hazard rate model. After running a series of Cox regressions we found out that the first default factor – *Original LTV* – has no statistically significant explanatory power (see Appendix for actual results). Therefore we will remove it from our model and from now on we will use only 3 factors for modeling the hazard rate of time to default. Let us name the 3 remaining covariates as follows:

$X_{DTI,i}$	Original Debt-to-Income ratio of a borrower i
$X_{UN,i}$	Unemployment rate of the quarter preceding default event of borrower i
$X_{IR,i}$	Euribor 3-month Interest Rate (monthly average of the month preceding default event of borrower i)

Table 5.1 Covariates

Recall that in our model the hazard rate for the distribution of mortgage duration (lifetime) is given by the following:

$$\lambda(t \mid \mathbf{X}_i) = h(t) * \exp(\beta' \mathbf{X}_i)$$

To obtain estimates of the beta coefficients and the baseline hazard function, we use the build-in Cox proportional hazards regression function – *'coxphfit'*, which is included in the Statistical Toolbox® of Matlab®.

The maximum likelihood estimation of the beta coefficients has produced the following results:

Covariate:	DTI - $X_{DTI,i}$	Unemployment - $X_{UN,i}$	Euribor 3m - $X_{IR,i}$
Beta coefficient	2.792984442	37.66940855	60.52762111
p-value	8.02754E-49	2.00197E-84	1.0386E-148
standard error	0.190191281	1.934816002	2.330536474
z-statistics	14.68513399	19.46924592	25.97153994

Table 5.2 Coefficient estimates

And the Cumulative Hazard Function $H(t) = \int_0^t h(s)\, ds$ for the distribution of mortgage duration has the following shape:
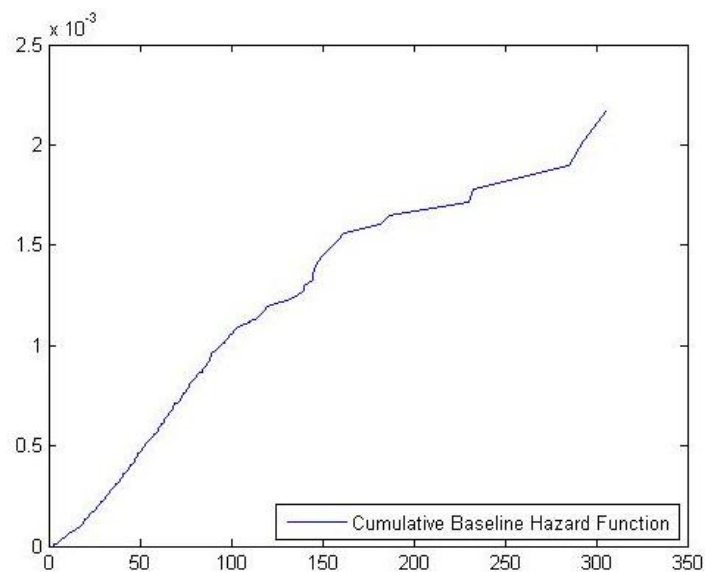
Figure 5.3 Baseline Cumulative Hazard

We can see that the cumulative baseline hazard function has very low values even for high durations. This of course is what we expected since mortgage default is a very rare event. Our estimate of the cumulative baseline hazard function is only given for durations less than or equal to 305 months, which is the

maximal duration of a defaulted loan in our database. This does not constitute a flaw in the model since in practice we almost never have to analyze mortgage contracts that have been outstanding for more than 25 years (300 months).

Example:

Let us now use our hazard rate model to compute the probability that a specific mortgage contract will default in the next year (i.e. the expected 12-month PD). Consider a mortgage loan l with original DTI ratio – 30% that has been issued on 01/07/2006 and suppose that the current quarterly unemployment rate in the Netherlands is 4.8% and the current monthly average of the 3-month Euribor interest rate is 1.5%. In other words we have: $X_{DTI,l} = 30\%$, $X_{UN,l} = 4.8\%$, $X_{IR,l} = 1.5\%$. Then according to (3.4) the cumulative hazard function for the distribution of the lifetime for this specific loan is:

$$\Lambda(t \mid \mathbf{X}_l) = \exp(\beta' \mathbf{X}_l) H(t) = \exp(\beta_1 * 0.3 + \beta_2 * 0.048 + \beta_3 * 0.015) * H(t)$$

and the β's are given in table 5.2.

Moreover, by definition 2.1, the cumulative distribution function for the duration of l is given by: $F_l(t) = P(D_l \leq t) = 1 - \exp\left(-\Lambda(t \mid \mathbf{X}_l)\right)$ and it has the following shape:
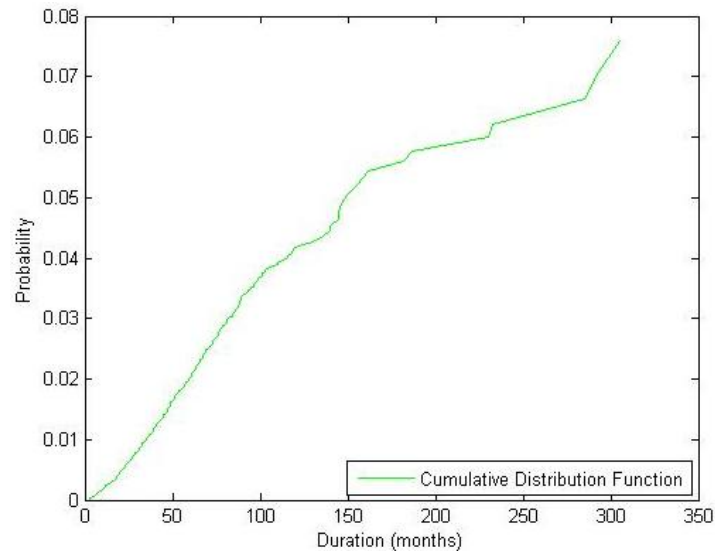
Figure 5.4 Cumulative Distribution Function

Note that the shape is very similar to the shape of the cumulative baseline hazard function. This is because of the multiplicative relationship between baseline hazard $H(t)$ and specific hazard $\Lambda(t\,|\,\mathbf{X}_l)$ and the fact that $1-\exp(-x) \approx x$ for small x.

Now since $\theta_l = 37$ (months since date of issue) and following (3.2), the expected 12-month probability of default for mortgage contract l is:

$$P(\tau_l \leq 12) = P(D_l \leq 49\,|\,D_l > 37) = \frac{P(D_l \leq 49) - P(D_l \leq 37)}{P(D_l > 37)} = \frac{0.01603 - 0.01071}{1 - 0.01071} = 0.005377$$

5.2 Expected Loss and Loss Distribution

The expected loss due to default of a mortgage contract i can be characterized by the following:

$$\mathrm{E}(Loss) = PD * \mathrm{E}(LDG) + (1 - PD) * 0 = PD * \mathrm{E}(LGD) \qquad (5.1)$$

and for the mortgage contract l that we considered in the previous example, we get the following expected loss:

$$\mathrm{E}(Loss) = 0.005377 * (1 - 0.8959) = 0.056\%$$

As we mentioned before, from risk management's point of view, not only the expected loss is important but also the uncertainty around it – in other words we are more interested in the whole probability distribution of the expected loss.

Since in our model the PD depends on two uncertain in the future factors, namely unemployment and interest rates, the stochastic distribution of the expected PD will be determined by the stochastic distribution of those factors. Or put in other words, we can derive the distribution of the expected PD by simulating a large number of possible realizations for unemployment and interest rates. In the same way we can use the beta distribution that we explained and estimated in section 4.4, to simulate a number of LGD realizations. Combining PD and LGD simulations, we are able to derive the whole distribution of the loss arising from a default of a mortgage contract (or a portfolio of mortgage contracts as we will see in the next section).

The simulated probability distributions of PD, LGD as well as the distribution of the expected Loss for this specific mortgage contract are presented in the Appendix (figures **A11**, **A12** and **A13**).

5.3   Scenario Simulations

To simulate interest rate paths we use a technique widely used in mathematical finance. We assume that the evolution of interest rates is a mean reverting process that can be described by the Cox-Ingersoll-Ross (CIR) model (*Cox, Ingersoll and Ross* [17]).   The model specifies that the short term interest rate follows the following stochastic differential equation:

$$dr_t = a(b - r_t)dt + \sigma\sqrt{r_t}\,dW_t \qquad\qquad (5.2)$$

where a is the mean reversion parameter, b is the long-term mean (equilibrium level), σ is the volatility and W_t is a standard Brownian motion. The CIR model is an extension of the well-known Vasicek model (*Vasicek, Oldrich* [18]).

Simulating unemployment rates is a bit trickier since in the literature there are no classical models describing unemployment evolution as a stochastic random variable. There are, however, a number of models that use ARMA or ARCH regressions to forecast unemployment rates. We have to make it clear here, that in this paper we are not interested in forecasting the most likely future unemployment rate, but in all possible realizations of unemployment rates, i.e. we are interested in the whole stochastic distribution of unemployment. That is why, to generate our theoretical unemployment realizations, we use the historical distribution of the relative change of unemployment (quarterly) for the last 9 years. We have determined (see Appendix) that for this time period the distribution of the relative change in unemployment is best fitted by a *normal* distribution with mean

$\mu_{un\_r} = 0.00193$ and standard deviation $\sigma_{un\_r} = 0.04034$ respectively. The idea is to model a possible evolution of unemployment in the following way:

$$un_{t+1} = un_t * (1 + un\_r_t) \qquad (5.3)$$

where un_t is the level of unemployment at time t and $un\_r \sim N(\mu_{un\_r}, \sigma_{un\_r})$ is a normally-distributed random variable, with the above mean and standard deviation. un_r_t represents the relative change in unemployment from time t to time $t+1$.

We will run our simulations with 3 basic unemployment scenarios. The first one we will call *standard*. In this scenario we assume that the monthly relative change of unemployment will keep its historical mean and standard deviation and the future evolution of unemployment will be described by (5.3) with the historical mean $\mu_{un\_r} = 0.00193$ and standard deviation $\sigma_{un\_r} = 0.04034$. In this scenario the evolution of unemployment is described by paths that are almost evenly distributed around the current level of unemployment and have a slight upward trend (mean is positive but small, see Appendix).

On the other hand, in the current credit crisis we expect that unemployment will rise (and it has been rising for some time already). Although we are not trying to forecast the most likely future unemployment we have to take into account that at this moment it is much more likely that unemployment will increase substantially in the near future. The Dutch Central Bank (De Nederlandsche Bank – DNB) has two stress testing scenarios for the unemployment rate in the Netherlands. The first one – *DNB base scenario* suggests that the unemployment rate will be around 5.5% at the end of 2009 and around 8.7% at the end of 2010. The second one –

the *DNB stress scenario* suggests that unemployment rate will be around 5.7% at the end of 2009 and around 9.7% at the end of 2010.

In order to simulate the evolution of unemployment that corresponds to these scenarios we simply adjust the distribution of the relative change in unemployment in the following way – we keep the monthly standard deviation the same, but we increase the mean of the normally distributed random variable un_r_t in (5.3). As a consequence we get simulated paths with much higher upward trend, compared to the *standard* scenario (see Appendix for actual results figures **A10**, **A11** and **A12**). These scenarios reflect better the current macroeconomic outlook and should produce more realistic expectations of the future PD's and expected losses for residential mortgages.

An overview of unemployment and interest rate simulated scenarios is presented in the Appendix (figures **A7** to **A12**).

5.4   Loss Distribution of RMBS Collateral Pool

The collateral pool of a typical RMBS transaction usually consists of thousands of mortgage contracts. In order to obtain the loss distribution for the whole mortgage portfolio we just sum the expected absolute losses of all loans in the pool and express the result as a percentage of the total outstanding balance of the pool. Again, by simulating a large number of single mortgage losses we are able to obtain the loss distribution of the whole collateral pool.

Example: Storm 2007-I B.V.

Storm 2007-I B.V. is a EUR2bn true sale securitization transaction of mortgage loans, originated in the Netherlands by Obvion N.V. The Collateral Pool this RMBS transaction can be briefly characterized by the following:

Original Principal Balance:	2 047 484 181 €
Number of Borrowers:	10 499
Average Loan Per Borrower	194 309 €
Weighted Average Seasoning (months)	6
Weighted Average Original LTV	83.7%
Weighted Average DTI	29.7%

Table 5.3 Collateral Pool

Unfortunately we do not have detailed information on each single loan contract and we will use the weighted averages in our PD model. This of course means that in our simulation all mortgages will have the same probability of default, which is not the case in reality. We hope that this lack of detailed information will not have a crucial impact on analyzing the RMBS transaction.

To obtain the loss distribution of the Pool, we first generate 5000 random realizations of unemployment and interest rates. We use those in our PD model to obtain 5000 realizations of a single mortgage contract probability of default. In each different simulation all contracts have the same PD. For each of those 5000 simulations, we simulate a random LGD value (as described in section 4.4) for each of the 10 499 loans in the pool. The sum of the absolute losses of all loans expressed as a percentage of the outstanding balance is then the expected loss of the pool. In this example we will simulate one year period and calculate the loss

distribution for 1 year. Let us first concentrate on the standard scenario for the simulation of unemployment. Simulations for other scenarios are presented in the Appendix. Figure 5.3 represents the expected loss distribution for the above collateral pool of mortgages.
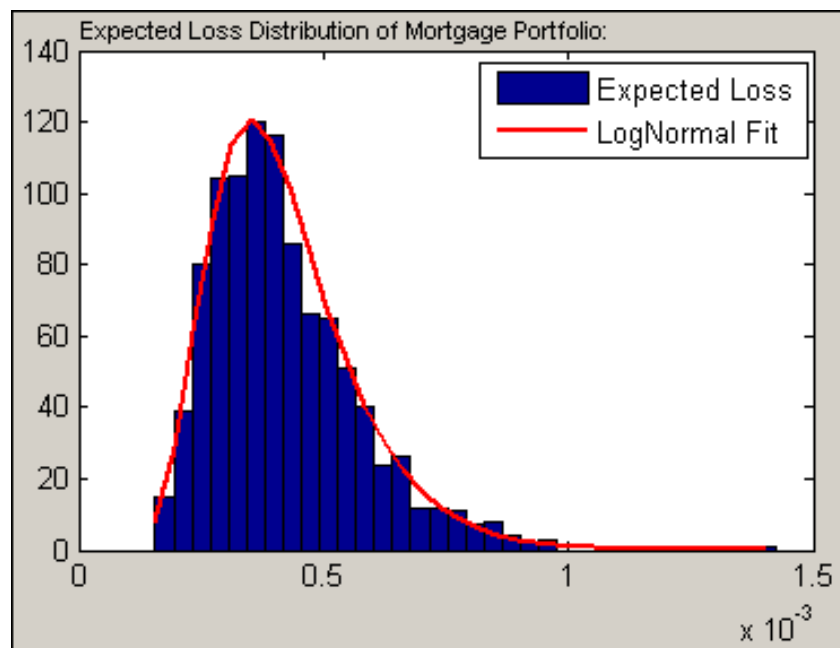
Figure 5.3 Expected Loss Distribution (Standard unemployment Scenario) – results for other scenarios are presented in the Appendix

The following table summarizes some important in risk management characteristics of the loss distribution:

Mean	0.0422%
Standard Deviation	0.0166%
95%-quantile	0.0735%
99%-quantile	0.0991%

Table 5.4 Loss distribution characteristics

5.5   Loss Distribution and Defaults of RMBS Notes

Let us again focus on Storm 2007-I B.V. The Notes structure of this RMBS transaction is the following:

Class	Rating	Size (%)	Size (EURm)	Credit Enhancements
A1	AAA	10.0	200	5.00%
A2	AAA	17.0	340	5.00%
A3	AAA	69.0	1 380	5.00%
B	AA	2.0	40	3.00%
C	A+	1.2	24	1.80%
D	A-	0.8	16	1.00%
E	BBB-	1.0	20	Excess Spread – 0.5%

Table 5.4 Notes Structure

The Credit Enhancement (CE) is linked to credit quality – it is a cushion that protects investors (notes holders) against losses that arise from the underlying pool.

In this RMBS transaction there are 3 types of credit enhancement. The first and most common is – subordination. Subordination means that a given tranche (class of notes) bears any losses only if the tranches junior to it have been fully exhausted. In other words losses are propagated from Class E to Class A1 notes. Subordination is one of the grounding principals of securitization. In this RMBS transaction subordination protects all the notes except Class E notes. The second type of credit enhancement is the Excess Spread (XS) – it protects the Class E notes. The XS is the difference between interest payments derived from the pool

of mortgages and the weighted average coupon paid on the notes. For this transaction the XS is guaranteed by a swap agreement and is fixed on 50 basis points per annum. The third credit enhancement is the reserve fund – this is a layer of protection that has to be exhausted before note holders bear any losses. At origination of the transaction the reserve fund is funded by the issuance of the E notes, i.e. the SPV issues an extra trance of equity notes (also sometimes called 'turbo notes' - in this case the E notes) to create a buffer of protection for the tranches senior to the equity notes.

We can think of the credit enhancements as the thresholds that need to be crossed so that certain class is affected by a loss arising from the collateral pool. Since losses from the pool are transferred to the notes from the most junior (in this case Class E notes) to the most senior (the A1 Class), a loss to a certain class of notes only occurs if the actual loss is big enough to fully exhaust all the tranches junior to it, i.e. if the loss crosses the credit enhancement threshold of this tranche.

Figure 5.4 represents the loss distribution of the notes for the Storm 2007-I B.V. RMBS transaction after a 5000 simulations (with the Standard scenario for unemployment) for a 1 year period. Here again we use the current unemployment of 4.8% and current 3month Euribor 1.5% as starting values of the scenario simulations.
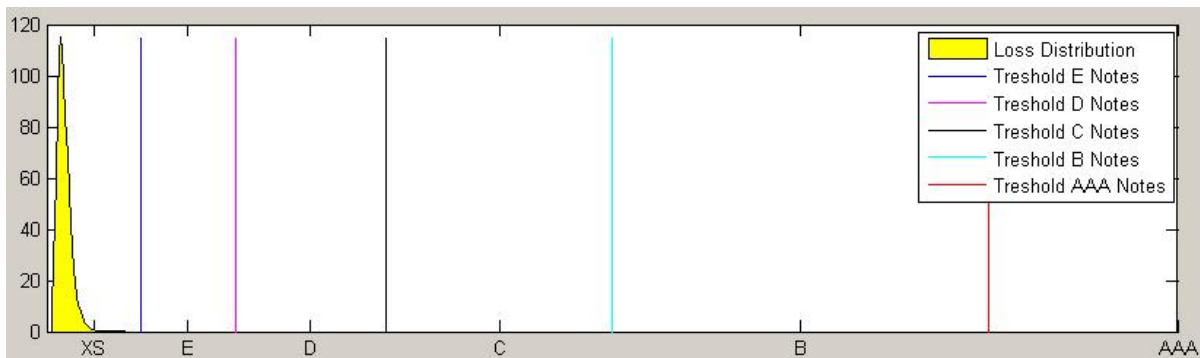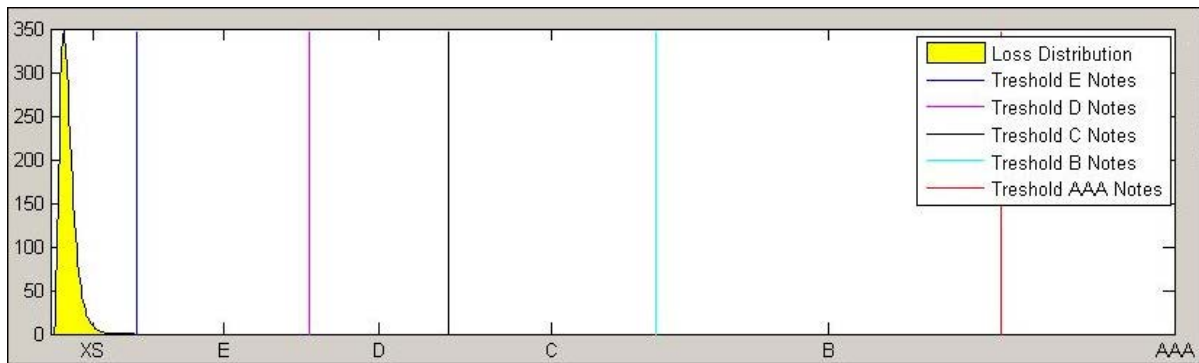
Figure 5.4 Loss Distribution of the Notes (1 year simulation period)

We can see from figure 5.4 that in this case there is no actual loss for the note holders. According to our model, all the losses arising from defaults of mortgages in the collateral pool are absorbed by the available excess spread.

Let us now concentrate on a situation where unemployment will rise rapidly in the near future – as predicted by the base and stress scenarios of the DNB.
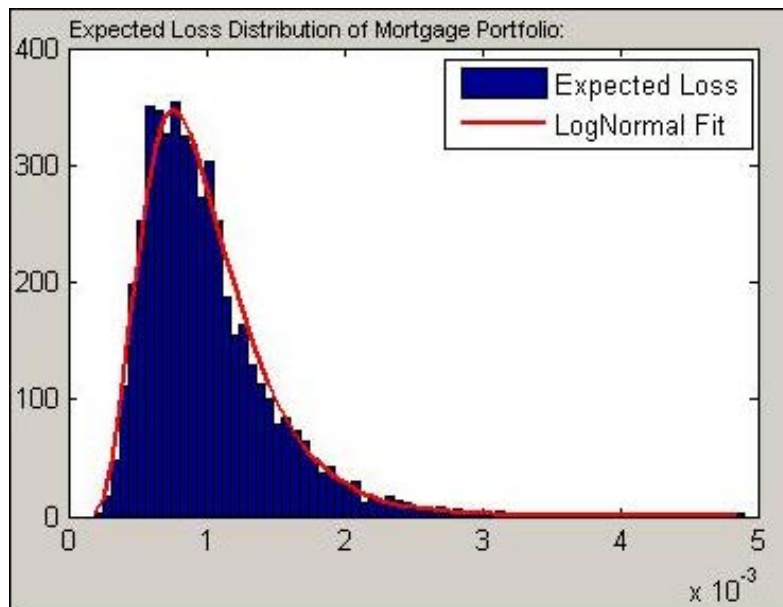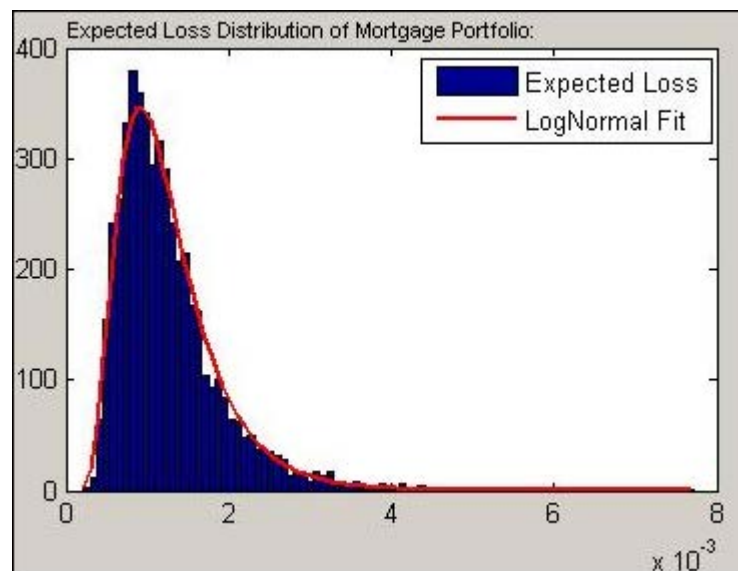

Figure 5.5 Expected Loss Distribution of the Collateral pool (DNB Base Scenario)

Figure 5.6 Expected Loss Distribution of the Notes (DNB Base Scenario)

As we can see from figure 5.6, according to our model, even if unemployment increases rapidly as predicted by the DNB base in scenario (to a level of 8.7% by the end of 2010), there will be no losses to the Notes of this particular RMBS deal. On the other hand in the DNB stress scenario there occur some losses to the E Notes but again those losses are relatively rare - only in 10 out of 1000 simulations the loss is big enough to affect class E Notes (see fig 5.8).

Figure 5.7 Expected Loss Distribution of the Collateral pool (DNB Stress Scenario)
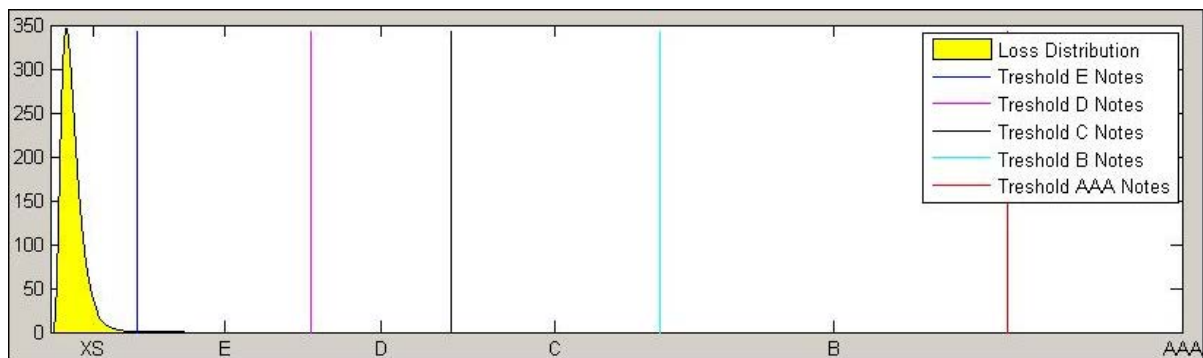
Figure 5.8 Expected Loss Distribution of the Notes (DNB Stress Scenario)

We have done all of the above simulations assuming that there will be no significant decline in the residential property market in the Netherlands, which will effect the proceeds from selling the collateral of defaulted loans. This so called *Market Value Decline* (MVD) will lower the recoveries from the collateral and increase the expected LGD and therefore will have an impact on the expected loss of the mortgage pool and the loss of the RMBS Notes. Fortunately our model allows us to take this possible market value decline into account in our simulations, simply by adjusting the α and β parameters of the beta distribution that we use for simulating recovery rates. For example we can simulate an extreme case where the DNB stress scenario for unemployment was to come true together with a MVD of 20% for a 1-year horizon. MVD of 20% simply means that recovery rates will fall with 20% on average. The loss distribution of the Collateral pool and the loss to the Notes for this extreme scenario are presented in figures 5.9 and 5.91, respectively.
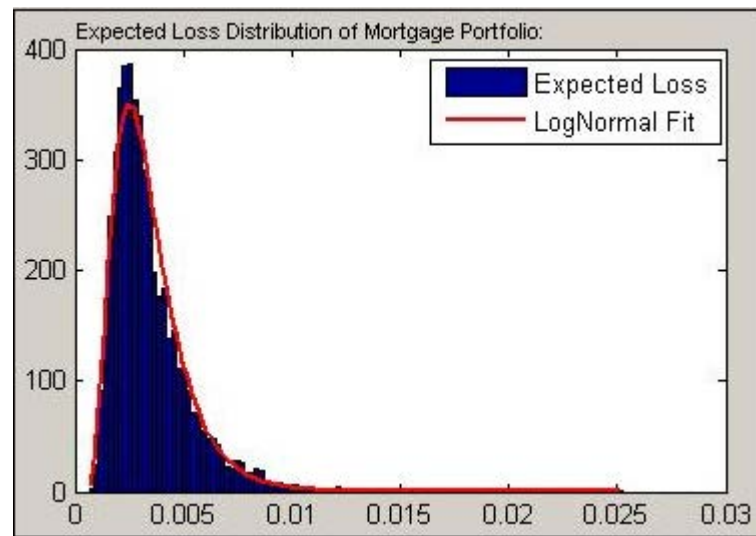
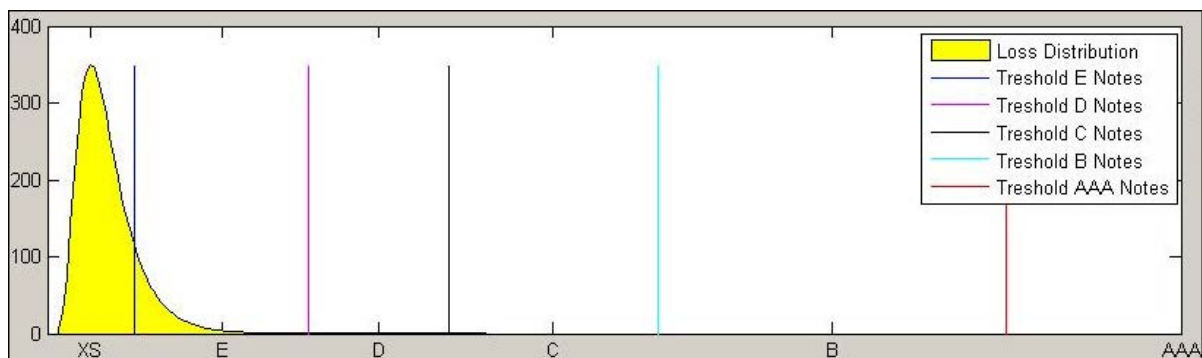**Figure 5.9 Expected Loss Distribution of the Collateral (DNB Stress Scenario with 20% MVD)**

Figure 5.91 Expected Loss Distribution of the Notes (DNB Stress Scenario with 20% MVD)

We can see here that in this extreme case there are significant number of simulations (767 out of 5000 to be exact) in which the expected loss is big enough to hit class E notes. There are even 5 cases in which the class D notes suffer losses and 1 case in which the C notes are hit. So if this scenario is a very likely scenario we could assign a probability of default $767/5000 = 15.34\%$ for the E notes

of this RMBS transaction, meaning that there is a 15.34% chance that E notes will suffer losses.

We have to mention again here that the goal of this project is not to identify the most likely development of future unemployment or interest rates or the future movements of the residential property market in the Netherlands. What is important for us here is that our model allows the user to input his own expectations of these variables and to adjust the analysis according to his own view of the macro economy. Our model presents a way to analyze the default risks associated with RMBS transactions for any possible evolution of unemployment, interest rates and residential property market.

Table 5 summarizes the results we have obtained after running a series of simulations with different assumptions about the evolutions of the stochastic factors in our model in 1-year horizon.

Notes Probability of Default				
	MVD	Standard	DNB Base	DNB Stress
A1 Notes	0%	0.00%	0.00%	0.00%
	10%	0.00%	0.00%	0.00%
	20%	0.00%	0.00%	0.00%
	30%	0.00%	0.00%	0.00%
A2 Notes	0%	0.00%	0.00%	0.00%
	10%	0.00%	0.00%	0.00%
	20%	0.00%	0.00%	0.00%
	30%	0.00%	0.00%	0.00%
A3 Notes	0%	0.00%	0.00%	0.00%
	10%	0.00%	0.00%	0.00%
	20%	0.00%	0.00%	0.00%
	30%	0.00%	0.00%	0.00%
B Notes	0%	0.00%	0.00%	0.00%
	10%	0.00%	0.00%	0.00%
	20%	0.00%	0.00%	0.00%
	30%	0.00%	0.00%	0.00%
C Notes	0%	0.00%	0.00%	0.00%
	10%	0.00%	0.00%	0.00%
	20%	0.00%	0.00%	0.02%
	30%	0.00%	0.00%	0.08%
D Notes	0%	0.00%	0.00%	0.00%
	10%	0.00%	0.00%	0.00%
	20%	0.00%	0.04%	0.10%
	30%	0.00%	0.12%	0.64%
E Notes	0%	0.00%	0.06%	0.20%
	10%	0.00%	1.12%	3.96%
	20%	0.02%	6.42%	15.34%
	30%	0.06%	16.86%	32.32%

Table 5 - Probability of default for Note's tranches in different scenarios

Our model also allows us to simulate time horizons different than 1-year, which is the standard in many credit risk management tools. Simulation results for time horizons of 2 and 3 years are presented in the Appendix (figures **A18** to **A21**). We only present the simulation results for the DNB base scenario of unemployment and 20 % market value decline because we believe this is the most likely one in the current economic situation. Of course other scenarios can be easily calculated. We note that losses for longer periods are bigger because the probability of default of mortgage contracts for a longer time horizon is higher and consequently the expected loss of the portfolio is higher.

VI.   Conclusion

In this paper we have presented an approach for modeling the distribution functions of the Probability of Default, Loss Given Default and consequently the Expected Loss for a mortgage contract or a portfolio of mortgage contracts. We have considered the distribution of mortgage lifetime as well as the distribution of time to default and the associated conditional intensity processes, given the set of predictors for the default event. We have modeled the intensity (hazard rate) of mortgage duration as a function of two macroeconomic covariates (unemployment and interest rates) and one mortgage-specific variable (debt-to-income ratio). Our hazard rate model turned out to be very suitable in translating the stochastic behavior of the macroeconomic variables to the behavior of expected and unexpected PD and the stochastic distribution of the Expected Loss for a portfolio of mortgage loans. Our model is flexible with respect to the choice of default predictors (as long as this choice is economically sound). The model is also very well suited for performing stress testing on one or all of the factors that influence the cash flows of RMBS tranches.

We have estimated the model using a non-parametric partial likelihood approach based on the Cox Proportional Hazards model, which enabled us to estimate the effects of default predictors to the default intensity process.

We have then run a series of simulations and determined the stochastic distribution of the Expected Loss of a collateral pool of mortgages and translated these expected losses to the Notes of the a typical RMBS transaction. This approach makes it possible to analyze the loss distribution of different RMBS tranches and gives insights of the default risks associated with RMBS

transactions. The model can be used for the analysis of the default risk of any Dutch RMBS. Although the model was designed to analyze RMBS transactions originated in the Netherlands, it can be easily adopted for other countries, provided that enough data on the history of residential mortgage defaults is available.

Further research has to be done in the direction of better handling missing covariate values. This could improve the estimation of the coefficients for the PD model and consequently - the sensitivity and significance of certain default predictors. Another possible improvement could be taking into account a hazard rate model that allows for time-dependent covariates, which would also solve the problem with missing covariate observations.

Other possible improvements to the model can include modeling the LGD in dependence of LTV. We know that the loan-to-value ratio has an impact on the recovery rates and can be used to better describe residential mortgage LGD. This can be done for example, by estimating a different beta distribution for predefined LTV buckets – but for reliable estimates a larger (than the one we had) database of LGD's is required. The dependence structure of LGD and PD on one hand, and unemployment and interest rates on the other, should also be investigated. Based on the data we use we have found no significant correlation between PD and LGD but this could be due to the fact that the time period that we base our estimates on (2002-2008) is a relatively stable period with no extreme economic shocks. A database that spans over a larger period of time and includes data from economic downturns, could improve the accuracy of our model.

Appendix

Coefficient	LTV	DTI	Unempl.	Euribor 3m
Beta	-7.4E-08	2.80631	53.477971	71.41378169
p-value	0.989268	3.01E-49	3.06E-86	6.1286E-179
st. erorr	5.51E-06	0.19024	2.7170636	2.503752991
z-statistics	-0.01345	14.75144	19.682267	28.52269451

Table A1 – Cox model coefficients statistics (with LTV)

Beta Covariance matrix			
3.04E-11	4.04E-10	6.53E-08	6.75915E-08
4.04E-10	0.036191	-0.040678	0.091384154
6.53E-08	-0.04068	7.3824345	2.378678034
6.76E-08	0.091384	2.378678	6.26877904

Table A2 – Covariance matrix of coefficient estimates (with LTV)

	Beta 1	Beta 2	Beta 3
Beta	2.8063	53.4781	71.4132
p-value	3.0137e-49	3.0377e-86	6.0560e-179
st. error	0.1902	2.7170	2.5037
z-statistic	14.7514	19.6826	28.5231

Table A3 – Cox Model Regression statistics (without LTV)

	1	2	3
1	0.0362	-0.0407	0.0914
2	-0.0407	7.3823	2.3785
3	0.0914	2.3785	6.2686

Table A4 – Covariance matrix of coefficient estimates (without LTV)

*Note that the coefficient estimate of LTV is practically 0 and its *p*-value is almost 1 (significantly different of 0) and also the parameter estimates for the other covariates are the same with or without LTV – i.e. LTV has no significant impact in modeling the hazard rate of default.
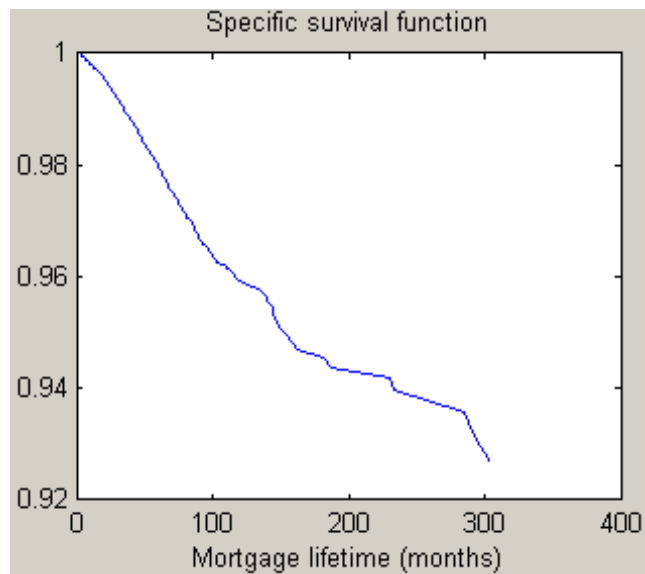
Figure A5 – Survival Function of mortgage contract l: $X_{DTI,l} = 30\%$, $X_{UN,l} = 4.8\%$, $X_{IR,l} = 1.5\%$.
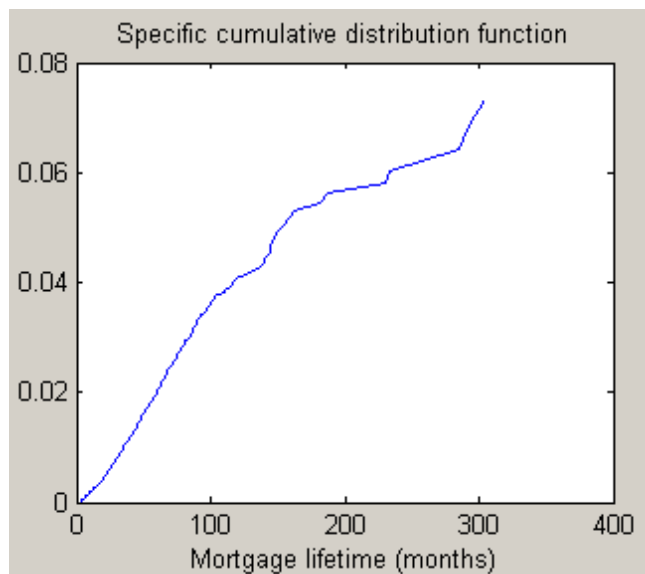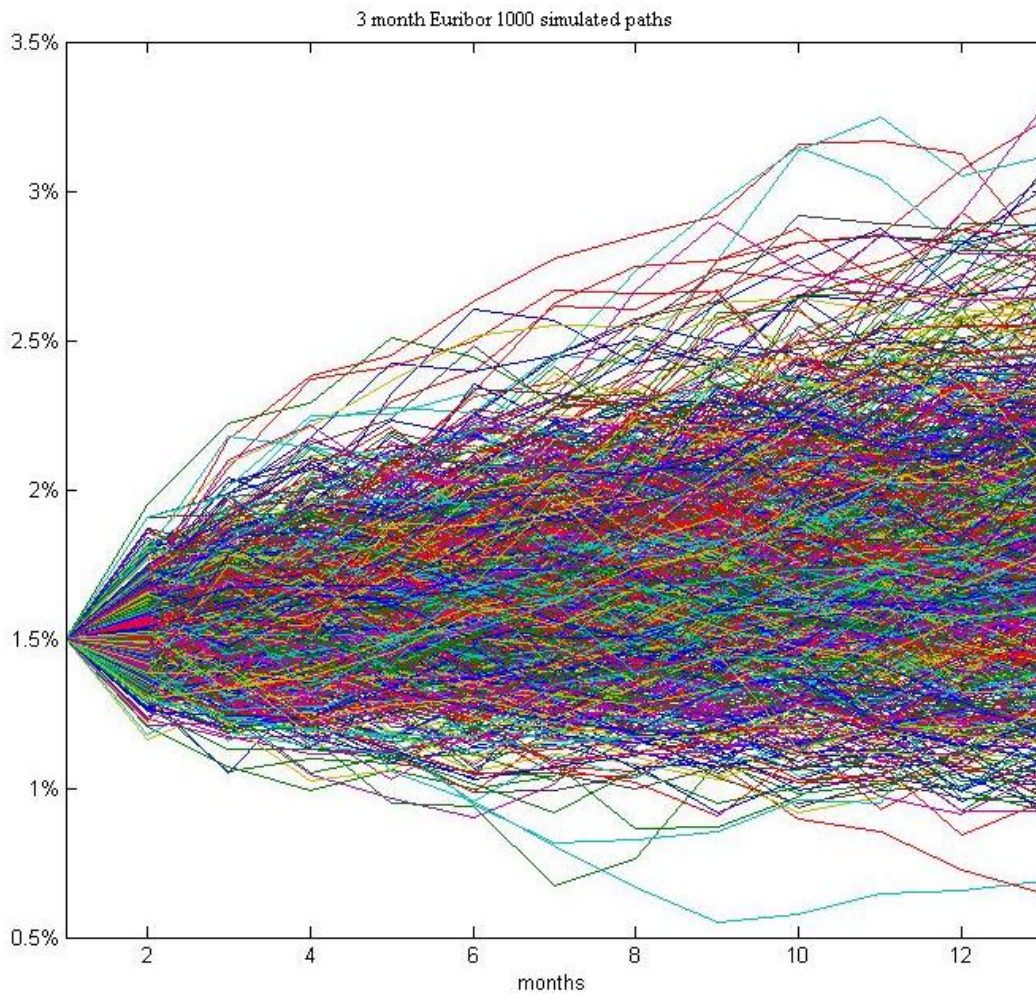

Figure A6 – Cumulative Distribution Function of mortgage contract l: $X_{DTI,l} = 30\%$, $X_{UN,l} = 4.8\%$, $X_{IR,l} = 1.5\%$.

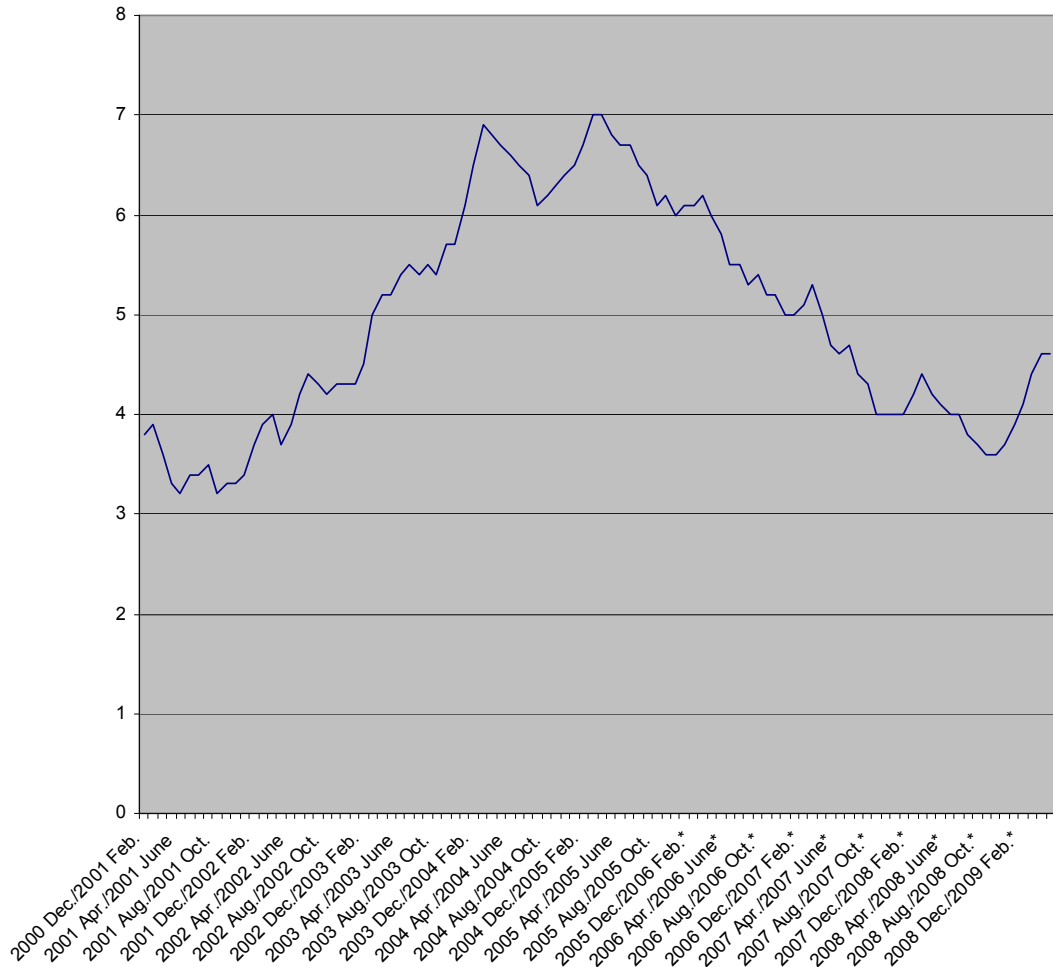**Figure A7 – Euribor 3month simulations – 1000 paths for 12 month period**

CIR Process:

$$dr_t = a(b - r_t)dt + \sigma\sqrt{r_t}\,dW_t$$

with parameters (estimated on monthly data from the last 9 years):

$a = 0.014849660801059$ (mean reversion parameter)

$b = 0.033987279169537$ (long term mean)

$\sigma = 0.010397560084086$ (volatility)

Figure A8 – Evolution of the quarterly unemployment rate in the Netherlands for the last 9 years.

Unemployment statistics:

Min	Max	Mean	Median	Std. Dev.
3.20%	7.00%	4.96%	4.85%	1.13%
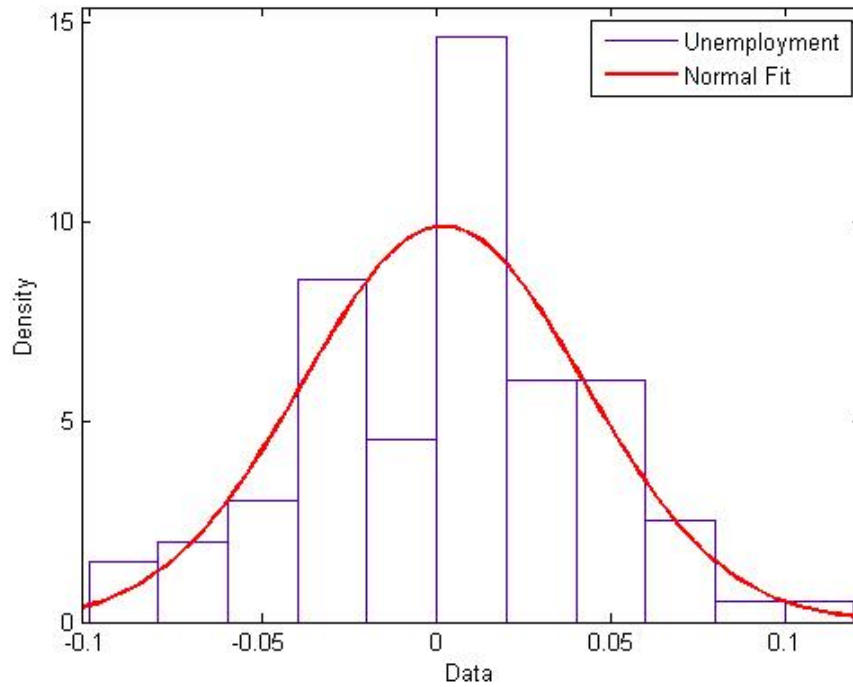
Figure A9 – Relative change in unemployment – Distribution and Normal fit.

Relative change in unemployment – historical statistics:

Mean	Std. Dev.	Median
0.00193	0.040335285	0

Normal fit – estimated parameters:

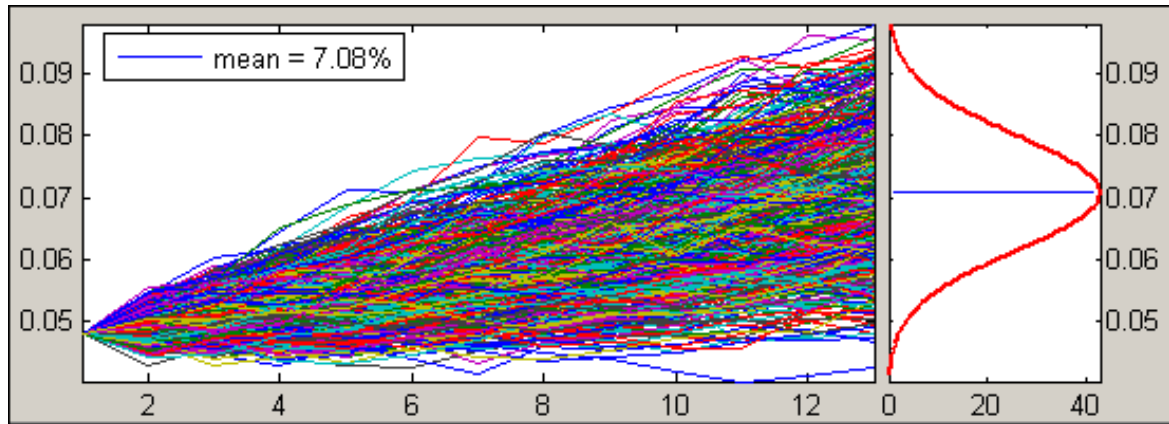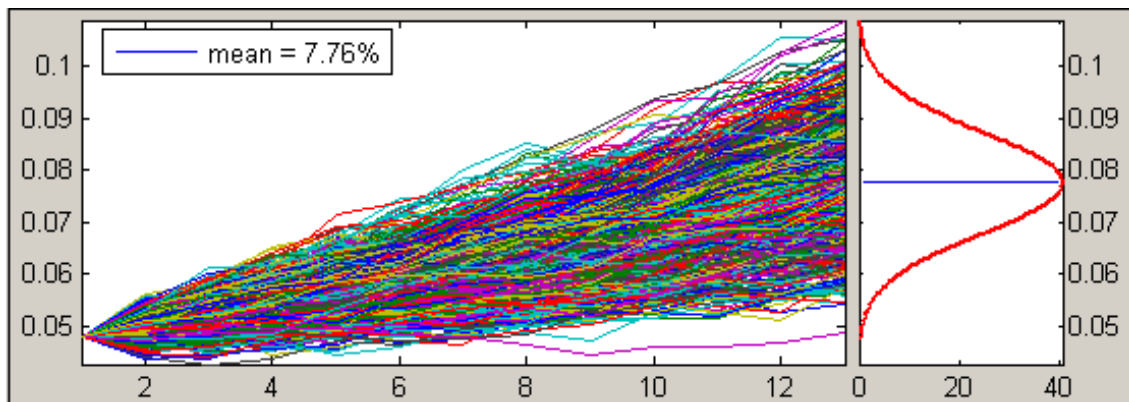| Mean	Std. Dev.
0.00192985	0.0403353
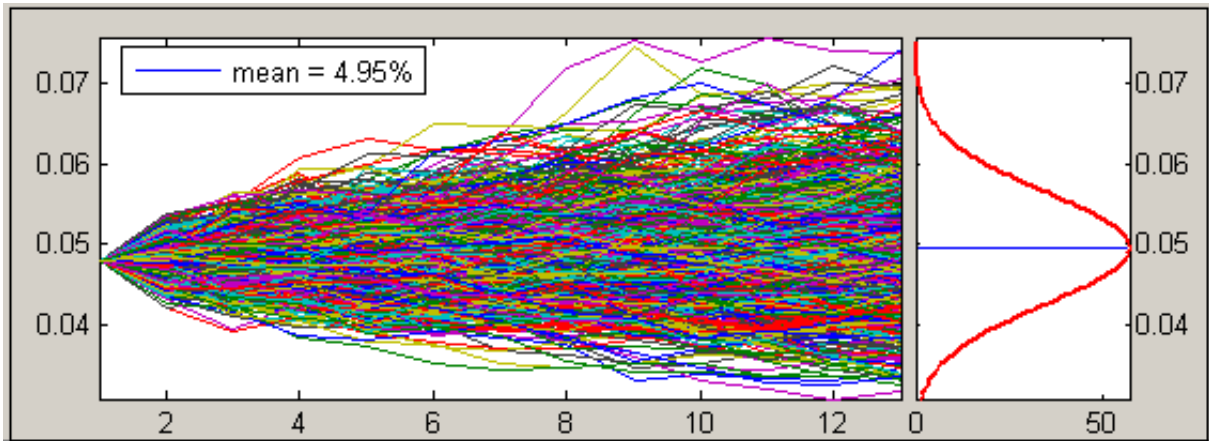
Figure A10 - Unemployment simulations – 1000 paths for 12 month period

Unemployment paths (Standard Scenario):

$$un_{t+1} = un_t * (1 + un\_r_t)$$

where $un\_r \sim N(\mu_{un\_r}, \sigma_{un\_r})$ is a random relative change following a normal-distribution with the following parameters (estimated on data for the last 9 years):

$\mu_{un\_r} = 0.00192985$ (mean)

$\sigma_{un\_r} = 0.0403353$ (standard deviation)

**Figure A11 - Unemployment simulations – 1000 paths for 12 month period**

Unemployment paths (DNB Base Scenario):

$$un_{t+1} = un_t * (1 + un\_r_t)$$

where $un\_r \sim N(\mu_{un\_r}, \sigma_{un\_r})$ is a random relative change following a normal-distribution with the following parameters (estimated on data for the last 9 years):

$\mu_{un\_r} = 0.0335$ (mean)

$\sigma_{un\_r} = 0.0403353$ (standard deviation)

Figure A12 - Unemployment simulations – 1000 paths for 12 month period

Unemployment paths (DNB Stress Scenario):

$$un_{t+1} = un_t * (1 + un\_r_t)$$

where $un\_r \sim N(\mu_{un\_r}, \sigma_{un\_r})$ is a random relative change following a normal-distribution with the following parameters (estimated on data for the last 9 years):

$\mu_{un\_r} = 0.0405$ (mean)

$\sigma_{un\_r} = 0.0403353$ (standard deviation)
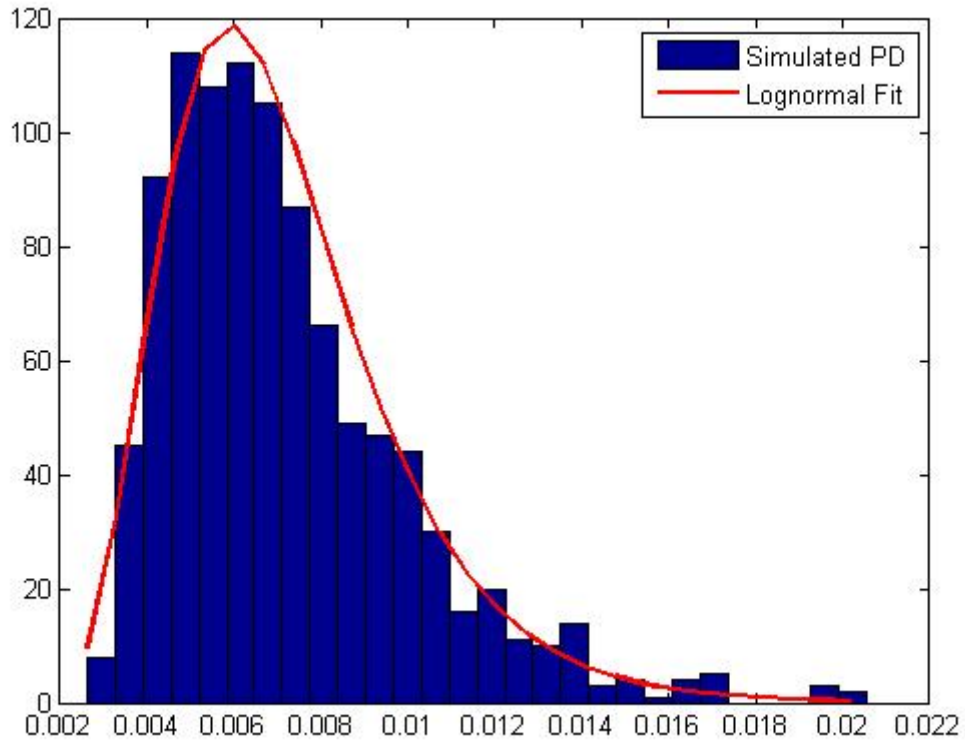
Figure A13 – Expected PD Distribution of mortgage contract l: $X_{DTI,l} = 30\%$

1000 simulations with initial $X_{UN,l} = 4.81\%$ **and initial** $X_{IR,l} = 1.5\%$
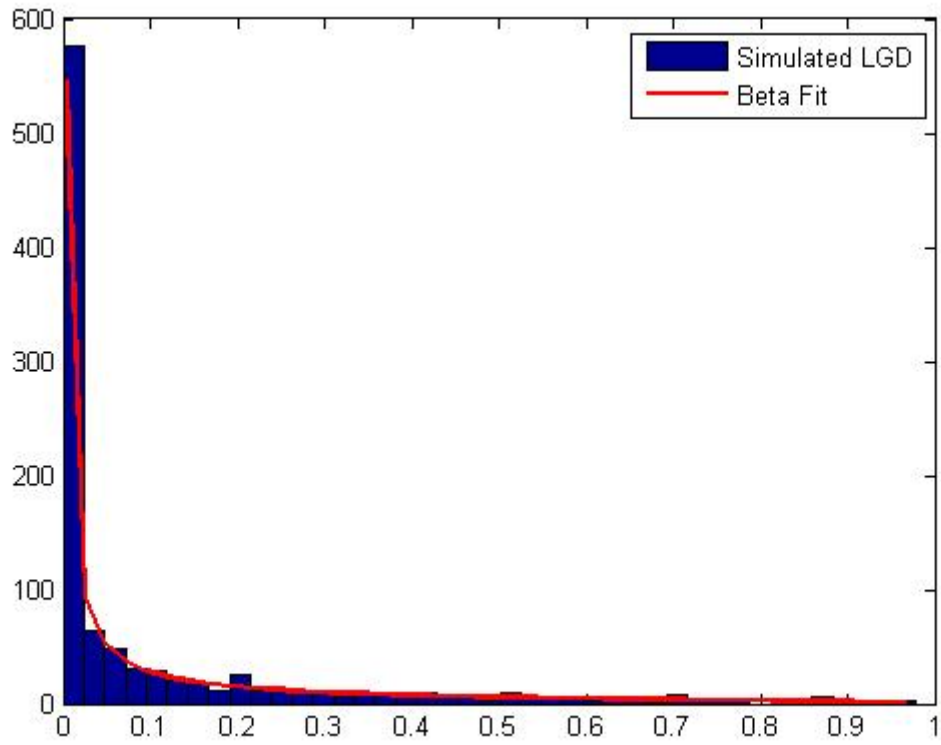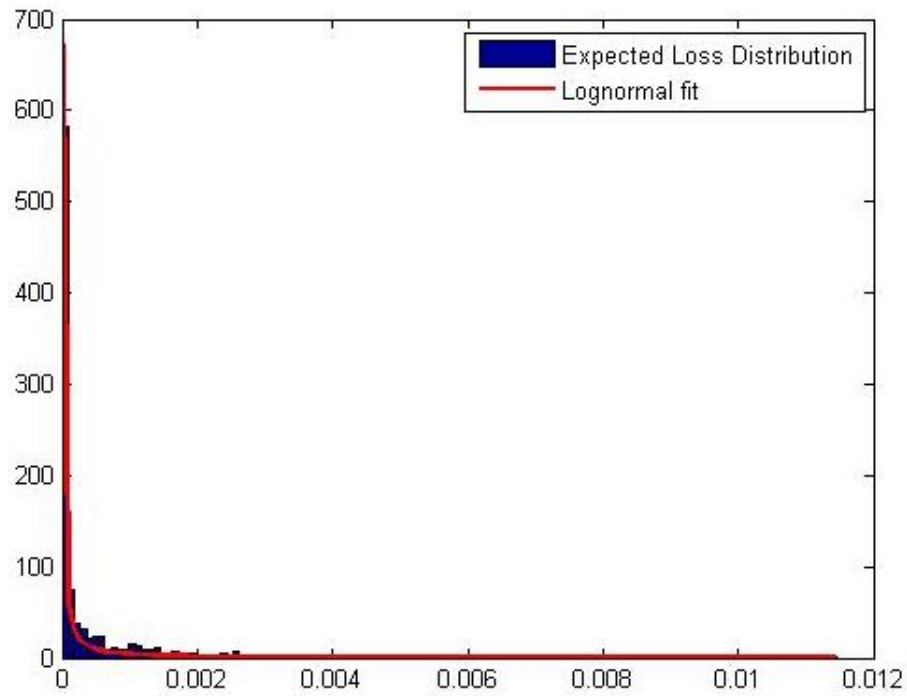
Figure A14 – Expected LGD Distribution of mortgage contract l: $X_{DTI,l} = 30\%$

1000 simulations with initial $X_{UN,l} = 4.81\%$ **and initial** $X_{IR,l} = 1.5\%$

Figure A15 – Expected Loss Distribution of mortgage contract l: $X_{DTI,l} = 30\%$

1000 simulations with initial $X_{UN,l} = 4.81\%$ **and initial** $X_{IR,l} = 1.5\%$

Note that even though PD is not that low the Expected Losses are very small because only ≈ 50% of the defaults result in actual loss

Collateral

Pool characteristics

|			
Original principal balance (EUR)	2,047,484,181	Regional concentration (%)	
Current principal balance (EUR)	2,040,050,722	Zuid-Holland	15.6
Average current loan per borrower (EUR)	194,309	Noord-Brabant	22.4
		Noord-Holland	12.4
Number of borrowers	10,499		
Seasoning (months)	6		
Loan to Value (LTMV) (%)		Lien position	
WA OLTMV (%)	83.7	First and subsequent ranking (%)	100
WA CLTMV (%)	83.5		
WA Indexed CLTMV (%)	82.6	Jumbo (%)	10
Mortgage characteristics (%)		Payments	
Interest only	69.7	Payment frequency	Monthly
Savings	6.3	Payment method	Direct Debit
Insurance	16	Performing loans (%)	100
Repayment	0.1	WA DTI (%)	29.7
Investment	7.0		
Floating rate loans	5.4		
Fixed rate loans	94.6		
WA Interest rate (%)	4.37		
Interest index (Euribor)	3m		

Source: Transaction documents

Figure A16 - STORM 2007-I B.V. Collateral pool characteristics.

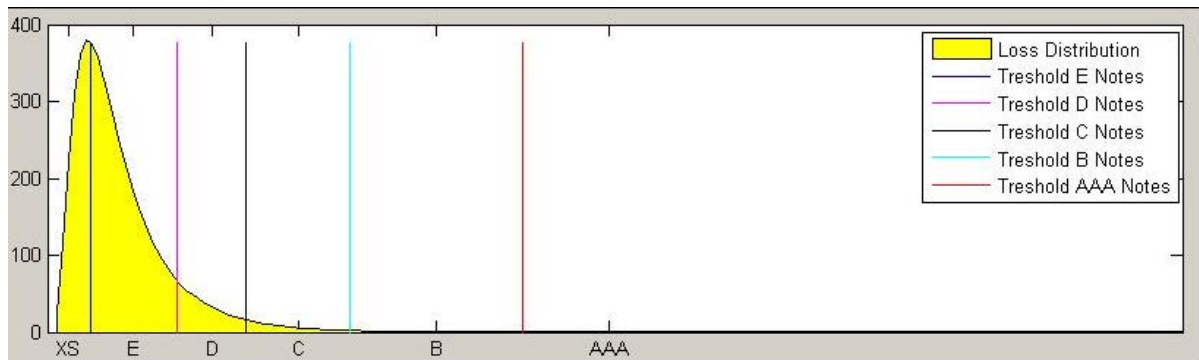## Capital Structure

| Class	Rating	Size (%)	Size (EURm)	CE (%)
A1	AAA	10.0	200	5.00
A2	AAA	17.0	340	5.00
A3	AAA	69.0	1,380	5.00
B	AA	2.0	40	3.00
C	A+	1.2	24	1.80
D	A-	0.8	16	1.00
E	BBB-	1.0	20	

Figure A17 - STORM 2007-I B.V. Notes Structure
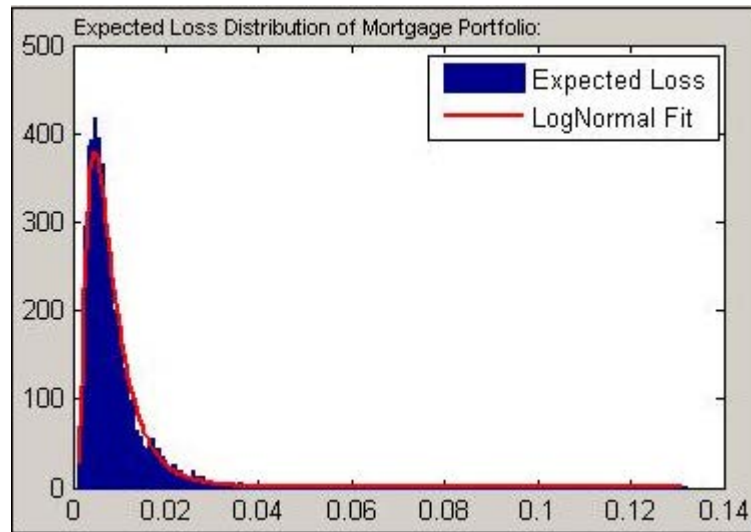
Figure A18 - Expected Loss Distribution of Collateral pool (DNB Base Scenario with MVD 20%) – 2 year simulation period - 5000 simulations
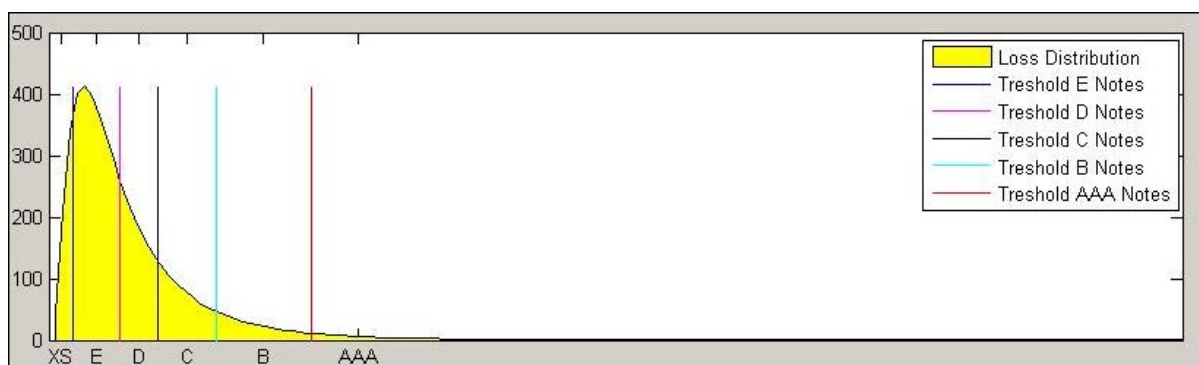

Figure A19 - Expected Loss Distribution of the Notes (DNB Base Scenario with MVD 20%) –

2 year simulation period - 5000 simulations
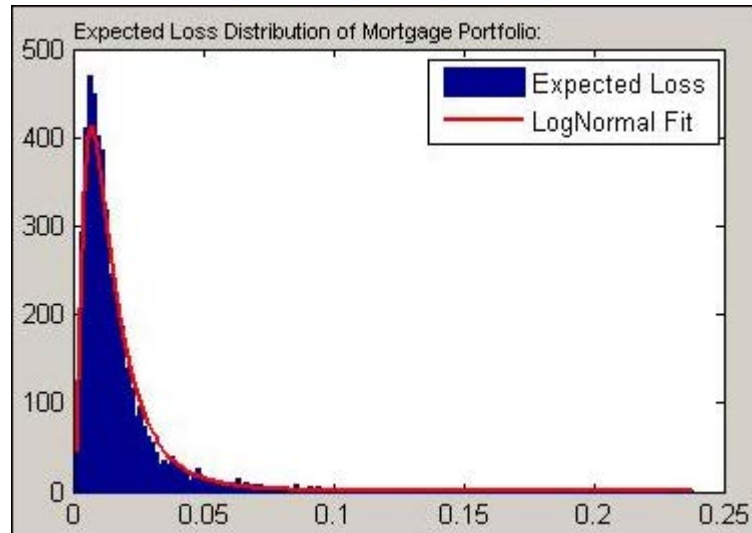
Figure A20 - Expected Loss Distribution of Collateral pool (DNB Base Scenario with MVD 20%) – 3 year simulation period - 5000 simulations

Figure A21 - Expected Loss Distribution of the Notes (DNB Base Scenario with MVD 20%) –

3 year simulation period - 5000 simulations

References

[1] McNeil A.J., Frey R. & P. Embrechts: *"Quantitative Risk Management: Concepts, Techniques and Tools", Princeton University Press, 2005*

[2] Cox D.R. (1972): "Regression Models and Life Tables", *Journal of the Royal Statistical Society, Series B, 34:187-220*

[3] Deng, Y. (1997): "Mortgage Termination: an Empirical Hazard Model with Stochastic Term Structure", *Journal of Real Estate Finance and Economics, 14(3), 309-331*

[4] Deng, Y., Quigley, J.M. & R. Van Order (2000): "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options", *Econometrica 68, pp.2775-307*

[5] Merton, R.C. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", *Journal of Finance, 29(2):449-470*

[6] Titman S. & W.N. Torous (1989): "Valuing Commercial Mortgages: An Empirical Investigation of the Contingent-Claims Approach to Pricing Risky Debt", *Journal of Finance, 44(2): 345-373*

[7] Kau J.B., D.C. Keenan, W.J. Muller, & J.F. Epperson (1992): "A Generalized Valuation Model for Fixed-rate Residential Mortgages", *Journal of Money, Credit and Banking 24(3), 279-299*

[8] Kau J.B. & D.C. Keenan (1998): "Patterns of Rational Default", *working paper, University of Georgia.*

[9] Smith L., S. Sanchez & E. Lawrence (1999): "A Comprehensive Model for Managing Credit Risk on Home Mortgage Portfolios", *Decision Sciences, 27(2), 291-317*

[10] Santos Silva J. & J. Murteira (2000): "Estimation of Default Probabilities Using Incomplete Contracts Data", *Draft Paper, Universidade de Coimbra*

[11] Follian J. W.-V. Huang & J. Ondrich (1999): "Stay, Pay or Walk Away: A Hazard Rate Analysis of FHA-insured Mortgage Terminations" *Draft Paper, Freddie Mac and University of Syracuse*

[12] Deng, Y & J. Quigley (2000): "Woodhead Behaviour and the Pricing of Residential Mortgages", *working paper W00-004, Institute of Business and Economics Research, University of California ant Berkeley*

[13] De Giorgi E. (2001): "An Intensity Based Non-Parametric Default Model for Residential Mortgage Portfolios", RiskLab report, available at "http://www.risklab.ch"

[14] Cox D. & D. Oakes (1984): "Analysis of Survival Data", *Chapman and Hall, London*

[15] Cox D. (1975): "Partial Likelihood", *Biometrika 62:269-276*

[16] Breslow N.E. (1972): "Contribution to the Discussion of paper by D.R. Cox", *Journal of the Royal Statistical Society, Series B, 34:216-217*

[17] Cox J.C., J.E. Ingersoll & S.A. Ross (1985). "A Theory of the Term Structure of Interest Rates". *Econometrica 53: 385–407*

[18] Vasicek, Oldrich (1977). "An Equilibrium Characterisation of the Term Structure". *Journal of Financial Economics 5: 177–188*