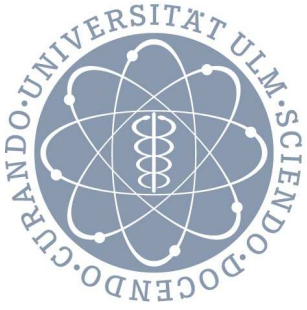
Modeling and Analysis of Structured Finance Products



Dissertation
zur Erlangung des Doktorgrades Dr. rer. nat.
der Fakultät für Mathematik und Wirtschaftswissenschaften
der Universität Ulm

vorgelegt von

Florian Kramer

aus
Kaufbeuren

Ulm, September 2008

Amtierender Dekan:	Prof. Dr. Frank Stehling
1. Gutachter:	Prof. Dr. Gunter Löffler
2. Gutachter:	Prof. Dr. Rüdiger Kiesel
Tag der Promotion:	12.12.2008

Contents

List of Symbols and Notation

We list all important expressions occurring in the text according to their order of first appearance. In order to discern between constants, stochastic processes and functions, we use the following notation:

Constants and sets are denoted by plain symbols, primary stochastic processes are followed by "(t)", whereas other stochastic processes as well as functions (which may map stochastic processes) are followed by brackets either explicitly showing their dependencies or just their dependence on time t.

In the text we will use vector- and matrix-notation throughout. This means that for a function $f : \mathbb{R}^2 \to \mathbb{R}^I$, $f(x,y) = (f_1(x,y), \ldots, f_I(x,y))^T$ denotes an I-dimensional columnvector and $f^T(x,y)$ its transpose. We write $f_{(\Pi)}(x,y)$ for the $|\Pi|$-dimensional sub-vector of $f(x,y)$ that contains all coordinates $f_i(x,y)$, $i \in \Pi \subseteq \{1,\ldots,I\}$, and $f_{\setminus i}(x,y)$ denotes the vector which consists of all components of the original vector $f(x,y)$ except the i-th component. Comparisons between vectors f and constants $c \in \mathbb{R}^I$ are to be understood component-wise, e.g. $f > c$ corresponds to $f_i > c_i$ for all $i \in \{1,\ldots,I\}$; if c is one-dimensional, $f > c$ corresponds to $f_i > c$ for all $i \in \{1,\ldots,I\}$.

Chapter 2

$\mathbf{F} = (\boldsymbol{\mathcal{F}_t})_{\mathbf{0 \le t \le T^*}}$	General model filtration
$\mathbf{F^X} = \left(\boldsymbol{\mathcal{F}_t^X}\right)_{\mathbf{0 \le t \le T^*}}$	Subfiltration generated by a process X
$\boldsymbol{\tau}$	I-dimensional vector of $\mathbf{F}$-stopping times
$\boldsymbol{I}$	Dimension of τ
$\boldsymbol{N(t)}$	I-dimensional indicator process associated with τ
$\boldsymbol{E}$	I-dimensional vector of mutually independent, $Exp(1)$-distributed random variables

$\boldsymbol{\Lambda(t)}$	I-dimensional jump-trigger process
$\boldsymbol{X^1(t)}$	d_1-dimensional background process
$\boldsymbol{X^2(t)}$	d_2-dimensional contagion process, adapted to $\mathbf{F}^N$
$\mathbb{R}_+$	Positive real numbers
$\mathbb{R}_{\mathbf{0}+}$	Positive real numbers including $\{0\}$
$\boldsymbol{\nu}$	M-dimensional vector of Lévy measures
$\boldsymbol{\|\varsigma\|}$	Euclidian norm of $\varsigma \in \mathbb{R}^d$
$\boldsymbol{P_i(t,T)}$	Survival probability of i over $[t,T]$
$\boldsymbol{Q_i(t,T)}$	Jump probability of N_i over $[t,T]$
$\boldsymbol{_{T,t}p_i}$	Conditional survival probability of i over $[t,T]$
$\boldsymbol{_{T,t}q_i}$	Conditional jump probability of N_i over $[t,T]$
$\boldsymbol{\Pi}$	Subset of $\{1,\ldots,I\}$, i.e. $\Pi \subseteq \{1,\ldots,I\}$
$\boldsymbol{P_\Pi(t,T)}$	Joint survival probability of Π over $[t,T]$
$\boldsymbol{Q_\Pi(t,T)}$	Joint jump probability of $N_{(\Pi)}$ over $[t,T]$
$\boldsymbol{_{T,t}p_\Pi}$	Conditional joint survival probability of Π over $[t,T]$
$\boldsymbol{_{T,t}q_\Pi}$	Conditional joint jump probability of $N_{(\Pi)}$ over $[t,T]$
$\boldsymbol{A_i(t)}$	Compensator of N_i
$\boldsymbol{1_{\tau_i>t}\lambda_i(t)}$	Intensity process of N_i
$\boldsymbol{\Sigma_\Pi(t)}$	Density process of the change of measure $\frac{d\mathbb{P}^\Pi}{d\mathbb{P}}$
$\boldsymbol{1_{\tau_{(\Pi)}>t}h_\Pi(t)}$	Intensity process of $1 - 1_{\tau_{(\Pi)}>t}$
$\boldsymbol{1_{\tau_{(\Pi)}>t}\lambda_\Pi(t)}$	Intensity process of $N_\Pi(t) := 1_{\forall\, i\neq j\in\Pi:\, \tau_i=\tau_j\leq t}$
$\boldsymbol{1_{\tau_{(\Pi)}>t}\lambda^*_\Pi(t)}$	Intensity process of $N^*_\Pi(t) := 1_{\forall i\neq j\in\Pi,\, l\notin\Pi:\, \tau_i=\tau_j\leq t,\, \tau_i\neq\tau_l}$
$\boldsymbol{_{T,t}F_\Pi}$	Dependence function of $N_{(\Pi)}$ over $[t,T]$
$\boldsymbol{_{T,t}\varphi_X}$	Laplace transform of the increments of a process X over $[t,T]$
$\boldsymbol{_{T,t}\rho_{ij}}$	Linear correlation between $N_i(T)$ and $N_j(T)$ given the information at $t<T$
$\boldsymbol{L(t)}$	The loss process
$\boldsymbol{A(t)}$	Compensator of the loss process
$\boldsymbol{H_k(t)}$	Process that counts jumps of the loss process L with size k
$\boldsymbol{\lambda^\perp_k(t)}$	Intensity process of H_k

$_{T,t}\boldsymbol{\sigma_L}$	Expected volatility of L in $[t, T]$
$\boldsymbol{Y^1(t)}$	The Ornstein-Uhlenbeck component of the affine process
$\boldsymbol{Y^2(t)}$	The non-negative component of the affine process
$\boldsymbol{Z(t)}$	Finite-state Markov process describing the regime of the system
$\boldsymbol{r}$	Dimension of the state space of Z
$\boldsymbol{\mathcal{P}(\Pi)}$	Power set of a set Π

Chapter 3

$\boldsymbol{R^1(t)}$	Process summarizing all macroeconomic variables
$\boldsymbol{R_i^2(t)}$	Process summarizing all firm-specific variables of firm i
$\mathbf{1}_{\tau_i > t}\boldsymbol{\lambda_i(t)}$	Default intensity of firm i
$\boldsymbol{M(t)}$	Time-changed portfolio loss process
$\boldsymbol{\mathcal{L}}$	Likelihood function
$\boldsymbol{\theta}$	Parameter vector
$\boldsymbol{\lambda^c(t)}$	Factor component of the intensities
$\tilde{\boldsymbol{\lambda}}_{\boldsymbol{i}}\boldsymbol{(t)}$	Idiosyncratic component of the intensities
$_{t+\Delta,t}\boldsymbol{g_X}$	Transition density of a process X over $[t, T]$
$_{t+\Delta,t}\tilde{\boldsymbol{g}}_{\boldsymbol{X}}$	Approximate transition density of a process X over $[t, T]$
$\boldsymbol{\lambda^P(t)}$	Portfolio intensity
$\boldsymbol{\mathcal{I}}$	Intermediate quantity in the EM algorithm

Chapter 4

$\boldsymbol{i_t}$	Combined mortality index (CMI) at the end of year t
$\boldsymbol{Y_1(t)}$	The Baseline Component in our model
$\boldsymbol{Y_2(t)}$	The Catastrophe Component in our model

List of Abbreviations

ABS	Asset Backed Securities
AR	Accuracy Ratio
a.s.	almost surely
AXA	AXA Cessions
BAJD	Basic Affine Jump Diffusion
BJD	Basic Jump Diffusion
bn	billion
bps	basis points
CAN	Canada
CAT	Catastrophe
CATM	Catastrophe Mortality
CDF	Cumulative Distribution Function
CDO	Collaterized Debt Obligation
CDS	Credit Default Swap
cf.	compare (confer)
CH	Switzerland
CMI	Combined Mortality Index
D	Germany
DCT	Dominated Convergence Theorem
e.g.	for example (exempli gratia)
EL	Expected Loss
EM	Expectation-maximization (algorithm)
etc.	and so forth (et cetera)
F	France
ff	and the following pages
FGIC	Financial Guarantee Insurance Co.
GSL	Gnu Scientific Library
I	Italy

i.e.	that is (id est)
J	Japan
Lehman	Lehman Brothers Inc.
LGD	Loss Given Default
Milliman	Milliman Inc.
ML	Maximum likelihood
mn	million
Moody's	Moody's Investors Service
ODE	Ordinary Differential Equation
OSIRIS	OSIRIS Capital Plc.
p.	Page
PD	Probability of Default
RCLL	right continuous with left limits
RMS	Risk Management Solutions
SAJD	Self-Affecting Affine Jump Diffusion
SAJDM	Self-Affecting Affine Jump Diffusion with Memory Effect
SALIC	Scottish Annuity & Life Insurance Company Ltd.
SDE	Stochastic Differential Equation
S&P	Standard and Poor's
Tartan	Tartan Capital Ltd. Series 1
U.K./UK	United Kingdom
U.S./US	United States of America
$	US Dollar
VBT	Valuation Basic Table
Vita I/II/III	Vita Capital I/II/III Ltd.
w.l.o.g.	Without loss of generality

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Formulation of the Problem

Capital markets have a long history: Merchants have traded financial securities long before the first bank was built on Wall Street – or even before Manhattan was acquired by Peter Minuit in 1626. For instance, the English word "bourse" as well as the German word "Börse" are said to have their roots in the name of a Dutch merchant family in Bruges called "van der Burse". In the 16th century, businessmen held meetings in front of the family's house for trading purposes. Since then, the world has seen a large number of financial innovations: Milestones are, for example, the initiation of a public trade with stocks in Amsterdam in the 17th century, or the issuance of the first traded government bonds in 1672 by the Dutch government.

In comparison with this long and eventful history of capital markets in general, the history of the financial innovation constituting the central topic of this thesis is relatively short: *Structured finance products* represent a young asset class with the first deals dating back to the 1970s. In spite of this brief history, they have already caused a great turmoil that reached its peak with the outbreak of a global financial crisis in 2007. The crisis has been accompanied by massive distortions of the financial system and dramatic losses of many market participants eventually resulting in an intervention of governments with the attempt to confine its impact. Still the crisis is present and there is an ongoing argument on the reasons leading to this financial disaster.

According to Fender and Mitchell (2005) p. 67, *"structured finance involves the pooling of assets and the subsequent sale to investors of tranched claims on the cash flows backed by these pools"*. In other words, cash flows of assets are repackaged into new securities, called *tranches*, that differ in their risk profile and are traded in capital markets. For example, a so-called Collateralized Debt Obligation

(CDO) tranche is a contingent claim on a debt portfolio. Cash flows from the debt portfolio are first used to serve the most senior tranche. Remaining funds are then distributed according to the tranches' rank in the seniority ladder. The market for such products has shown tremendous growth (see e.g. Duffie et al. (2007) or Fender and Mitchell (2005)) in the last years for various reasons: The issuance of structured securities, for instance, allows financial intermediaries to reduce their economic capital or realize arbitrage opportunities. On the other hand, investors can gain access to new asset classes such as mortality contingent securities. Moreover, the tranching allows for them to invest according to their risk appetite.

Due to this striking trade record, much research has been devoted to the topic of modeling the new securities. In particular, the number of CDO models has reached a vast number. Examples are the models of Duffie and Gârleanu (2001), Kalemanova et al. (2007), Papageorgiou and Sircar (2007), Graziano and Rogers (2006) or Herbertsson and Rootzén (2006). A common characteristic of structured finance is that the products usually reference a portfolio of securities where the payoff of each security is intimately linked to one particular event, and the challenge is to model the joint distribution of these events. For instance, a CDO tranche may reference a portfolio of corporate bonds, and its return strongly depends on the firms' default times.

Translated into mathematical terms, the primary task is to model a vector of *stopping times*, which correspond to the incidence of the relevant events in the underlying portfolio. In the scientific literature, there prevail two basic methodologies in order to approach the problem: The so-called *structural approach* and the *reduced-form* framework. For a detailed discussion, we refer to the books of Lando (2004) or Bielecki and Rutkowski (2002). Both approaches originate from the credit risk modeling world, but the latter – the reduced-form framework – has also been adopted to other fields of finance such as the modeling of mortality contingent securities (see Section 2.8).

In the structural approach, a firm's asset value process is modeled, and the default time is defined as the first time the asset value process falls below some given threshold. In most specifications of this model class, the default time can be predicted based on an interpretable factor, namely the firm's asset value. For examples of structural credit models, see e.g. the pioneering models of Merton (1974) and Black and Cox (1976) or – for a more recent contribution – the model of Kiesel and Scherer (2007). In the reduced-form framework, on the other hand, the default times are modeled as inaccessible stopping times by definition. Loosely speaking, this means that they come as a "surprise" and cannot be predicted. Model examples are Lando (1998) or Duffie and Gârleanu (2001).

Apart from this elementary difference, however, both approaches take advantage

of the same concepts in order to introduce inter-dependencies between the single default times: Common factors like the business cycle usually induce correlations in individual firms' default probabilities. Conditional on the stochastic evolution of common factors, defaults are independent, although some models additionally allow for the possibility that the failure of one company triggers other defaults, i.e. for so-called *contagion* effects.

In this thesis, we model default times as inaccessible stopping times and follow the reduced-form framework for several reasons. On the one hand, we will consider an application of our model framework in an insurance-related field and analyze securities depending on the future evolution of mortality rates (see Chapter 4). Similarities between reduced-form credit risk modeling on one hand and stochastic mortality modeling on the other hand are well-known and have been already pointed out by Artzner and Delbaen (1995) (see also Section 2.8), but it is far from obvious how to specify a stochastic mortality model within the structural approach. On the other hand, as pointed out by Duffie and Lando (2001), if the factors generating the stopping times are not perfectly observed a structural model automatically becomes a reduced-form model.

Our general model setup builds on the contributions by Lando (1998), Jarrow and Yu (2001) and Yu (2007). While Lando (1998) assumes that the realizations of the stopping times are independent conditional on some background driving process, in the *Primary-Secondary Framework* of Jarrow and Yu (2001) the realizations of a subset of stopping times (the primary stopping times) are allowed to influence the stopping times not belonging to this subset (the secondary stopping times) but not vice versa. Yu (2007) extends Jarrow and Yu (2001) and considers a setup with a background driving process, which is independent of the stopping times, but where the realizations of the stopping times can influence each other. Moreover, several authors have investigated model specifications with the possibility of simultaneous defaults (see e.g. Joshi and Stacey (2006)).

Our definition of the stopping times links both strands in the literature. It allows for conditional survival probabilities, i.e. the probabilities that the stopping times do not lie in a particular future time interval conditional on todays information, to depend on past realizations of the stopping times as well as on some background process which evolves independently of the stopping times. Furthermore, realizations of the stopping times may coincide. Hence, our final framework can be described as a Yu (2007) model extended by the possibility of simultaneous defaults.

This model setup builds the "platform" for our investigation of structured finance products. More precisely, we will focus on the following questions:

- *What types of models are needed to explain the characteristics of structured finance securities?*
- *Do different models imply similar profiles for the securities?*
- *When do simple and complex models lead to comparable results?*

Answering these questions requires a holistic approach. We tackle this problem from two directions: The first is a theoretical one. As already mentioned, in the recent past a heap of articles has been published all suggesting different models, but discussions in the papers are often limited to the respective model under consideration. Here, we neither focus only on one specific model, nor do we propose a particular model. We rather try to give a comparison of different models within our general model setup. The second direction from which we approach our basic questions is an empirical one: We investigate them in the context of real data. More precisely, when introducing new models we compare them with simpler, already established models from the literature whenever possible.

1.2 Outline and Contributions

This thesis is divided into three major chapters and two appendices. Following the introduction, Chapter 2 contains most of the theoretical results and builds the foundation for our applications in Chapters 3 and 4. In Chapter 3, we discuss the modeling and the risk analysis of structured credit products. Chapter 4 provides an investigation of mortality contingent catastrophe bonds.

In what follows, we provide a concise overview on the main contributions of this thesis and discuss the central results. A brief synopsis of the principal findings including final remarks can also be found at the end of each chapter.

As mentioned above, **Chapter 2** lays the groundwork for the applications of Chapters 3 and 4 and comprises most of the theoretical contributions. Here, we introduce and discuss a framework for modeling a vector of stopping times. Later in the applications, these stopping times will model the default times of firms in a credit portfolio or the insureds' times of death, respectively. What is special and novel about our disquisition is that it covers many models that have been proposed in literature and that are included in this framework as special cases.[1] Hence, our findings not only provide a deeper understanding of this type of models, but also help to structure the vast number of credit portfolio and CDO models presented

[1]Examples are Duffie and Gârleanu (2001), Lindskog and McNeil (2003), Mortensen (2006), Joshi and Stacey (2006), Graziano and Rogers (2006), Papageorgiou and Sircar (2007) and Chapovsky et al. (2006).

in the literature. For example, our investigations reveal that some specifications effectively imply similar or even almost identical models.

The chapter starts off with Section 2.1, in which the basic *stopping times model* considered throughout this thesis is introduced (Definitions 2.1.1 and 2.1.2). The model represents a flexible modeling "platform" that offers many appealing features: Conditional survival probabilities depend on some *background process*, which evolves independently of the stopping times, as well as on the past realizations of the stopping times. In addition, realizations of the stopping times can coincide. Speaking in terms of the credit portfolio application considered in Chapter 3, defaults in the portfolio can affect the default probabilities of the surviving firms and simultaneous defaults are possible.

In Section 2.2, we start our model analysis by initially focussing on one single stopping time. Proposition 2.2.2 derives the intensity of the one-jump process associated with this stopping time. As our discussion of this intensity process shows, jumps of the process "triggering" the stopping time are dispensable from a single stopping time perspective. Based on the intensity process, in Proposition 2.2.1 we provide an expression for the single survival probability in our setup. In addition, similarities between our model and the setup studied in Collin-Dufresne et al. (2004) in the case of one single stopping time are pointed out.

Subsequently, in Section 2.3, we derive joint survival probabilities and the corresponding intensities for the entire vector. We show that joint jumps of the components of the process "triggering" the stopping times play an important role for the dependence structure between the stopping times. More precisely, without such jumps realizations of the stopping times cannot coincide. In addition, we establish in equation (2.2) an interesting link between our setup and the *common Poisson shock models* studied in Lindskog and McNeil (2003). As an example, we show that the *Intensity Gamma* model introduced by Joshi and Stacey (2006) is essentially a common Poisson shock model (Example 2.3.1).

Section 2.4 provides a construction and simulation algorithm for our stopping times model (Algorithm 2.4.1). Since the stopping times may coincide in our setup, the algorithm presents an extension of an algorithm given by Yu (2007).

Adjacently, we analyze the model-implied dependence structure from a static point of view, meaning that we consider the dependence structure between the survival events over a fixed time horizon. We introduce the important notions of *conditional independence* and *contagion* and in Proposition 2.5.1 state conditions under which our stopping times model becomes a conditional independence setup. Given conditional independence, in Theorem 2.5.2 we show that the model-implied dependence

structure is the one implied by the copula function studied in Marshall and Olkin (1988). Our result implies that many time-continuous models proposed in scientific literature entail a well-known dependence structure.

In addition, we examine the clustering of the stopping times in time by investigating the dynamics of the *loss process*, which counts the occurrence of the stopping times. This presents a dynamic characterization of dependencies between the stopping times. First, in Corollary 2.6.1 we adapt a time-change result which is due to Meyer (1971) to our setting. This will be at the bottom of our statistical tests considered in Subsection 3.1.4. We then introduce a quantity, which we call the *expected volatility* of the loss process, and propose to consider this quantity as a measure for the clustering of the stopping times. We discuss the measure's properties and demonstrate its usefulness in the context of concrete model specifications in Subsection 2.6.2.

Afterwards, in Section 2.7 we address the issue of model tractability. As a first step, in Subsection 2.7.1 we state conditions under which the characteristic function of a stochastic process can be computed semi-analytically by solving a system of ordinary differential equations. This is then proven in Theorem 2.7.1. Although we allow for more flexible specifications than are possible within the so-called exponential affine model class considered in Duffie et al. (2000) and Duffie et al. (2003), we get a comparable, high degree of analytical tractability. We then discuss in Subsection 2.7.2 how our stopping times model has to be customized such that the general result can be applied to the computation of survival probabilities and the distribution of the loss process. Both represent quantities which are central to our applications in Chapters 3 and 4.

We conclude the second chapter by establishing a link between the general stopping times model and the more specific setups considered for the applications. Also, we demonstrate the usefulness of the developed theoretical tools for analyzing concrete model specifications. As an example, the credit portfolio model proposed by Duffie and Gârleanu (2001) is discussed.

In **Chapter 3**, we link theory and application. The key challenge when modeling structured credit products is to correctly model the underlying portfolio and, in particular, its dependence structure. For example, misspecified dependence structures affect the forecasted loss distribution of the structured credit product in a non-linear way. Therefore, the primary focus of this chapter is the question of which model specifications describe the dynamics of credit portfolios observed in real data best. The implications of our findings for structured credit modeling are discussed.

In the first section of the chapter, we estimate default intensities for a large sample

of US and non-US corporates. In contrast to Das et al. (2007), we show that estimated default intensities are able to account for the observed default clustering, although we estimate the intensities based on observable covariates such as the firms' Expected Default Frequency (EDF) and do not introduce additional contagion effects or "frailty" variables (Table 3.3). In addition, when examining the ability of our regression intensity model to rank firms according to their default likeliness we find that its predictive power is higher than that reported by Duffie et al. (2007) for their regression intensity model and a similar data set (Table 3.2).

Subsequently, in Section 3.2 we introduce a time-continuous model in order to model the joint dynamics of the intensities. The model includes other established models as special cases. In the simplest case, intensities follow a *Cox-Ingersoll-Ross* process (see Cox et al. (1985)). Based on our theoretical results from Chapter 2, we derive a formula which is important for the model implementation (Proposition 3.2.1) and analyze the dynamics of the different model versions. In general, the model represents a solid basis for our empirical investigation since the different nested model versions can easily be compared. In particular, we can analyze the question of whether and of when simpler models are sufficient to model the intensities.

Having introduced the model, we consider its calibration on a single firm basis in Section 3.3. After developing a calibration algorithm in Subsection 3.3.1, we compare the ability of the different model versions to explain the intensity of each single firm and examine the predictive power of the models. Especially for firms of bad credit worthiness, we find that models with intensity jumps are better able to model the intensity dynamics than purely diffusion-based models (Table 3.7). Concerning default prediction, however, we find that more complex models do not lead to better results (Table 3.8).

Since all versions of the introduced model rely on the assumption of conditionally independent defaults, in Section 3.4 we investigate the implications of this assumption in detail. We simulate default data based on a model in which past defaults affect the default intensities of the firms. Subsequently, we estimate wrong models – all based on the conditional independence assumption – as well as the original model, which has actually generated the data, and investigate the ability of each model to forecast the portfolio loss distribution. We find that estimation errors are by far more influential than errors related to the assumption of conditional independence (Table 3.9). Even though contagion effects played a dominant role in the data generating model, some of the conditional independence models "lead" to results that are similar to those obtained from the estimated true model (Table 3.10).

Towards the end of this chapter, we present the model estimation on a portfolio

basis. After introducing our estimation algorithm in Subsection 3.5.1, we conduct model inference and compare the estimated model versions with respect to their ability to explain the portfolio intensity. In a second step, using the estimated parameters we simulate paths of the portfolio loss process and the corresponding portfolio intensity. In this way, we can compute a time series of ratings for hypothetical, structured credit products referencing the portfolio and eventually obtain transition matrices for these products. This allows us to compare the different model versions with respect to these matrices. We find that based on our data set simple and complex models eventually imply similar risk profiles of structured credit products (Table 3.17). In particular, general quantities like the average default rate in the underlying portfolio have a much stronger effect on the results than the model choice.

Chapter 4 at last, presents the second application of our stopping times model to the analysis and pricing of mortality contingent catastrophe bonds.

Since we focus on concrete transactions in the investigation, the chapter starts out with a concise overview on the market for such securities in Section 4.1 and briefly describes the transactions to be analyzed.

This is followed by the introduction of our model for analyzing and pricing mortality contingent catastrophe bonds in Section 4.2. The model specification consists of two components: A *Baseline Component*, which models the "regular" random fluctuations of mortality over time and is driven by a diffusion, and a *Catastrophe Component* governed by a non-Gaussian Ornstein-Uhlenbeck process. To the best of our knowledge, our model presents the first fully-dynamic approach for modeling these securities proposed in literature. Our discussion of the model illustrates that the model offers many appealing features. In particular, survival probabilities can be determined analytically up to the solution of ordinary differential equations as shown in Proposition 4.2.1, and – on this basis – the "classical actuarial toolbox" can be used to compute insurance premiums or benefit reserves.

Model inference is conducted in Section 4.3. There, we introduce and discuss in detail three different estimation procedures for the model: First, in Subsection 4.3.1, we consider the calibration of the model based on historical mortality data. We find that particularly the parameters of the catastrophe component are subject to high uncertainties. In a second step, in Subsections 4.3.3 and 4.3.2 we derive risk-adjusted parametrizations based on insurance prices and market quotes of catastrophe mortality bonds, respectively.

Finally, in Section 4.4 we compute loss profiles and spread levels for the considered securities using the estimated model parametrizations from Section 4.3. Our results are then compared to risk profiles provided by so-called *risk modeling firms*,

which present the primary basis for rating agencies' and investors' decisions. In general, we find that the profiles are subject to high uncertainties regarding the underlying data and should therefore be treated precautiously by all stakeholders. In particular, risk measures provided by the risk modeling firms are substantially lower than our results for all considered parametrizations although there are no structural differences in the outcomes, which indicates that the parametrizations used by the risk modeling firms are rather "optimistic" (Table 4.9). In addition, by analyzing the obtained risk-adjusted parametrizations we are eventually able to give an explanation for the fast growth of the mortality contingent catastrophe bond market in recent years.

Appendix A states a closed-form expression for an important transform of a so-called *basic-affine jump diffusion* process, which is due to Duffie and Gârleanu (2001). We also derive an exact simulation algorithm (Algorithm A.0.1) for this model. The algorithm extends the exact simulation algorithm for a special case of the model, a Cox-Ingersoll-Ross process, that can be found in Glasserman (2004), p. 124. Aside from this algorithm, the appendix contains further implementation details for the models of Chapter 3.

In **Appendix B**, we show how an approximate distribution of the underlying *combined mortality index* can be derived in the model of Chapter 4. Additional figures containing parameter sensitivities of this model are presented, too.

Chapter 2

A Model for a Vector of Stopping Times

Many structured finance products reference a portfolio of securities where the payoff of each security is intimately linked to one particular event. Consider for example a portfolio of term-life insurance contracts. It is self-evident that the insureds' death times are the most important risk driver of the portfolio. Another example would be a portfolio of corporate bonds; its return will strongly depend on the joint distribution of the firms' default times. To analyze securities that are contingent on the evolution of such portfolios, one needs to model the event times determining the risk profile of the underlying portfolio.

By introducing such a model, the current chapter builds the foundation for our analysis of structured finance products in Chapters 3 and 4. Considered from a mathematical point of view, the problem is the following: Specify a model for a vector of stopping times (associated with the incidence of the relevant events such as defaults), which provides enough flexibility to incorporate important phenomena observed in reality and whose parameters remain interpretable. We define the stopping times in a very intuitive way as

$$\tau_i := \inf \{t : \Lambda_i(t) \geq E_i\},$$

where the E_i are standard-exponentially distributed, mutually independent random variables, and Λ_i is a non-decreasing RCLL (right-continuous with left limits) process modulated by some background process X^1 as well as by the realizations of the τ_is themselves. Our specification covers many models that have been proposed in the literature as special cases. Consequently, the results of our model analysis provide a deeper understanding of these models and, in particular, help to structure the vast number of credit portfolio and CDO models which have been presented in the recent past.

The remainder of this chapter is organized as follows. After defining the stopping

times in Section 2.1 and introducing the basic quantities which will be important throughout the chapter, Section 2.2 shows how single survival probabilities can be calculated within the considered framework. In particular, we derive the intensities of the stopping times. In Section 2.3, we obtain joint survival probabilities and investigate the conditions under which realizations of the stopping times can coincide. A construction of the stopping times model is provided in Section 2.4; it represents a natural simulation algorithm for our model. Section 2.5 explores the model-implied dependence structure: In 2.5.1, we state conditions under which the considered setup eventually becomes a conditional independence setup and in Subsection 2.5.2 show that under certain conditions the model-implied dependence structure between the stopping times over a fixed time horizon can be identified with a well-known family of copula functions. The process counting the stopping times is investigated in Section 2.6: We present a time change result in Subsection 2.6.1, which will be at the bottom of our statistical tests conducted in 3.1.4, and in Subsection 2.6.2 propose a new measure to quantify the volatility of the aggregated process, i.e. a measure of the clustering of the stopping times in time. Section 2.7 addresses the problem of model tractability. In Subsection 2.7.1, we state conditions under which an important transform, often encountered in applications, can be calculated semi-analytically. In 2.7.2, we then show how this result can be applied to the computation of survival probabilities and the distribution of the aggregated loss process. Section 2.8 subsequently establishes the link between the general stopping times model and the more specific setups considered for our applications in Chapter 3 and Chapter 4. Furthermore, we demonstrate how the developed theoretical tools can be applied to analyze concrete model specifications. Section 2.9 concludes.

2.1 Definition of the Stopping Times Model

In this thesis, all stochastic processes and random variables are defined on a filtered probability space $\left(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{0 \leq t \leq T^*}, \mathbb{P}\right)$ where $\mathbf{F}$ models the flow of information and where $\mathcal{F}_{T^*} = \mathcal{F}$; $\mathbf{F}$ is further assumed to satisfy the usual conditions of $\mathbb{P}$-completeness and right-continuity. The basic object of the following investigation is an I-dimensional vector τ of $\mathbf{F}$-stopping times $\tau = (\tau_1, \ldots, \tau_I)^T$ and its corresponding indicator process[1]

$$N(t) = (1_{\tau_1 \leq t}, \ldots, 1_{\tau_I \leq t})^T .$$

Our basic *stopping times model* is defined as follows:

[1] A non-negative random variable τ_i is called an $\mathbf{F}$-stopping time if $\{\tau_i \leq t\} \in \mathcal{F}_t$ for every $t \geq 0$ (see e.g. Definition I.1 of Protter (2005)).

Definition 2.1.1 *Each τ_i is defined as*

$$\tau_i := \inf\{t : \Lambda_i(t) \geq E_i\},$$

where $E = (E_1, \ldots, E_I)^T$ denotes a vector of unit-exponentially distributed, mutually independent random variables and the jump-trigger process *$\Lambda = (\Lambda(t))_{0 \leq t \leq T^*}$ is defined as an **F**-adapted, I-dimensional, non-negative, strictly increasing process starting at zero, i.e. $\Lambda(0) = 0$, which is assumed to be RCLL. We assume that Λ can be written as*

$$\Lambda(t) = \int_0^t b'(s)ds + J(t),$$

with some integrable positive process b' and a pure jump process J showing only positive jumps. In addition, we are given

- *an **F**-adapted, d_1-dimensional real-valued* background process *$X^1 = (X^1(t))_{0 \leq t \leq T^*}$,*
- *an **F**-adapted, d_2-dimensional real-valued* contagion process *$X^2 = (X^2(t))_{0 \leq t \leq T^*}$ as well as*
- *positive, continuous functions*

$$\begin{aligned} b^1(\cdot,\cdot,\cdot): &\quad \mathbb{R}_{0+} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^I_{0+} \\ b^2(\cdot,\cdot,\cdot): &\quad \mathbb{R}_{0+} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^M_{0+} \end{aligned}$$

and

- *an M-dimensional vector ν of possibly time-dependent Lévy measures, concentrated on $\mathbb{R}^I_{0+}$ such that $\nu_m(t, \Xi) < \infty$ if Ξ is Borel on $\mathbb{R}^I_{0+}$ and bounded away from 0, $\nu_m(\{0\}) = 0$ and*

$$\int_{|\varsigma| \leq 1} |\varsigma|^2 \, \nu_m(t, d\varsigma) \; < \; \infty$$

for each $t \geq 0$ and $1 \leq m \leq M$.[2]

Then, the characteristics $(B(t), C(t), \vartheta(dt, d\varsigma))$ of Λ are given as

$$\begin{aligned} B(t) &= \int_0^t b^1(s, X^1(s), X^2(s))\, ds \\ C(t) &\equiv 0 \\ \vartheta(dt, d\varsigma) &= \nu\left(t, d\varsigma\right)^T b^2(t, X^1(t), X^2(t))\, dt. \end{aligned}$$

[2]For details, see e.g. Theorem I.44 of Protter (2005).

Remark 2.1.1 *As a consequence of our definition, the jump-trigger process Λ consists of a continuous part of finite variation and a jump part, i.e. the Brownian component is given by $C(t) \equiv 0$. b^1 is formulated throughout the text with respect to the truncation function $\chi(\varsigma) := 1_{|\varsigma|\leq 1}$.*[3]

As usual, when considering a particular model setup, the information flow plays a crucial role. To model this flow, we introduce the following subfiltrations of $\mathbf{F}$:

Definition 2.1.2

- $\boldsymbol{F}^{X^1} = \left(\mathcal{F}^{X^1}_{0\leq t\leq T^*}\right)$ *is defined as the filtration generated by the background process X^1, that is*[4]

$$\mathcal{F}^{X^1}_t = \sigma\left\{X^1(s),\, 0 \leq s \leq t\right\}.$$

- $\boldsymbol{F}^{\Lambda} = \left(\mathcal{F}^{\Lambda}_{0\leq t\leq T^*}\right)$ *is defined as the filtration generated by the jump-trigger process Λ, that is*

$$\mathcal{F}^{\Lambda}_t = \sigma\left\{\Lambda(s),\, 0 \leq s \leq t\right\}.$$

- $\boldsymbol{F}^{N_i} = \left(\mathcal{F}^{N_i}_{0\leq t\leq T^*}\right)$ *is defined as the filtration generated by N_i, i.e.*

$$\mathcal{F}^{N_i}_t = \sigma\left\{N_i(s),\, 0 \leq s \leq t\right\}.$$

  and $\boldsymbol{F}^N$ is generated by all coordinates of N, i.e. $\mathcal{F}^N_t = \mathcal{F}^{N_1}_t \vee \ldots \vee \mathcal{F}^{N_I}_t$. We assume that the contagion process X^2 is adapted to $\boldsymbol{F}^N$ and that $\mathcal{F}^{X^1}_{T^*}$ and $\mathcal{F}^N_t$ are independent conditional on $\mathcal{F}^{X^1}_t$.

- *The general filtration $\boldsymbol{F}$ is collectively generated by the processes Λ, X^1, and N, i.e.*

$$\mathcal{F}_t = \mathcal{F}^{X^1}_t \vee \mathcal{F}^{\Lambda}_t \vee \mathcal{F}^N_t.$$

Definition 2.1.2 ensures that the background process X^1 can essentially be considered as an exogenous process since the evolution of X^1 does not depend on the jumps of N. On the other hand, being adapted to $\mathbf{F}^N$, the contagion process X^2 reflects the influence of past jumps on the N_i. At this stage, it is important to note that – so far – we have only defined a model framework. However, it is not obvious whether the defined object is well-defined. As pointed out by Bielecki and Rutkowski (2002), this is a question which is not trivial in the case where past jumps of N_i are allowed to influence the jump probabilities of other N_j. We defer

[3] For a discussion of different truncation functions in the context of Lévy processes, see e.g. Cont and Tankov (2004), p. 83.

[4] We assume all filtrations to be augmented by the null sets of $\mathbb{P}$.

this question to Section 2.4, where we provide a construction of the introduced setup.

After the introduction of our model setup, it is worth discussing the possible roles that the processes which were involved in the definition of τ could play in later applications. As an illustration, we consider the *Intensity Gamma* model, which has been proposed by Joshi and Stacey (2006) to model the default times τ in a credit portfolio; it fits well into our framework:

Example 2.1.1 Intensity Gamma *(Joshi and Stacey (2006)). In this model, the default times τ are defined as in Definition 2.1.1 with*

$$\Lambda_i(t) = \sum_{j=1}^{p} \sum_{n=1}^{N-1} a_{in} \left(J_j(t_n) - J_j(t_{n-1}) \right) + \sum_{j=1}^{p} a_{iN} \left(J_j(t) - J_j(t_{N-1}) \right), \quad t \in [t_{N-1}, t_N],$$

where $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N < \ldots$ and the a_{in} are positive constants. J denotes a p-dimensional Gamma process with mutually independent coordinates; that is each coordinate of J is characterized through the Lévy measure $\vartheta_j(d\varsigma_j) = 1_{\varsigma_j > 0} \frac{c_j e^{-d_j \varsigma_j}}{\varsigma_j} d\varsigma_j$ ($c_j, d_j > 0$) and evolves independently of $E = (E_1, \ldots, E_I)^T$, too.[5] *In terms of Definition 2.1.1, we therefore have for all $s \geq 0$:*

$$\begin{aligned} 0 &\equiv X^1(s) = X^2(s) \\ 0 &\equiv b^1(s, X^1(s), X^2(s)) \\ 1 &\equiv b^2(s, X^1(s), X^2(s)) \qquad \textit{and} \\ \nu(s, d\varsigma) &= \sum_{j=1}^{p} \frac{c_j e^{-\frac{d_j}{a_{1n}} \varsigma_1}}{\varsigma_1} 1_{\varsigma_1 = \frac{a_{1n}}{a_{2n}} \varsigma_2 = \ldots = \frac{a_{1n}}{a_{In}} \varsigma_I > 0} \, d\varsigma, \qquad s \in [t_{n-1}, t_n]. \end{aligned}$$

In Example 2.1.1 as well as in our general setup, the riskiness of each bond intuitively depends on how fast the single coordinates of the jump-trigger process Λ increase: If we have a high probability that Λ_i grows very fast, the likelihood of the bond to survive will be low. In the current example, this likelihood depends on two quantities: First, it depends on the specification of J. Second, it depends on the constants a_{in} which determine how strong each Λ_i is exposed to J. Consequently, the a_{in} reflect different levels of default probabilities among the portfolio objects, while the specification of J is important for the dependence structure between the default times τ_i: Jumps of the J_j represent "shock events" which all objects are exposed to.

It is important to note that the Intensity Gamma model does – by far – not exhaust the full flexibility of the stopping times model introduced in Definition 2.1.1. The

[5] For further properties of a Gamma process we refer to Cont and Tankov (2004), Section 4.4.2.

example model does not make use of the background process X^1 that could, for example, model general macroeconomic conditions such as the current level of interest rates, on which default probabilities might depend. Generally, the background process may include all variables having an impact on default probabilities but that are themselves not influenced by defaults. Furthermore, the possibility of allowing past defaults to influence the default probabilities of the remaining, non-defaulted objects is not exploited since $0 \equiv b^1(s, X^1(s), X^2(s))$ and $1 \equiv b^2(s, X^1(s), X^2(s))$. It is worth mentioning that credit portfolio models in which defaults have an influence on the default probabilities of the surviving firms are typically called *contagion* models explaining the name of the contagion process X^2.

When analyzing a particular model, one is usually interested in the single or joint *survival* probabilities associated with the τ_i, which are the probabilities that the corresponding coordinates of N do not jump over a time interval $[t, T]$ given the information up to time t. We introduce these quantities next:

Definition 2.1.3 *1. $P_i(t,T)$ is defined as the probability of the event $\{\tau_i > T\}$ given the information until time t, i.e.*

$$P_i(t,T) := \mathbb{P}\left(\tau_i > T \middle| \mathcal{F}_t\right) = \mathbb{E}\left[1 - N_i(T) \middle| \mathcal{F}_t\right] = \mathbb{E}\left[1_{\tau_i > T} \middle| \mathcal{F}_t\right],$$

and $Q_i(t,T) := \mathbb{E}\left[1_{t<\tau_i \leq T} \middle| \mathcal{F}_t\right]$. The corresponding probabilities conditional on $\{\tau_i > t\}$ are denoted by

$$_{T,t}p_i := \mathbb{P}\left(\tau_i > T \middle| \mathcal{F}_t \wedge \{\tau_i > t\}\right)$$

and $_{T,t}q_i := \mathbb{P}\left(t < \tau_i \leq T \middle| \mathcal{F}_t \wedge \{\tau_i > t\}\right)$.

2. $P_\Pi(t,T)$, $\Pi \subseteq \{1, \ldots, I\}$, is defined as the probability of the event $\bigcap_{i\in\Pi}\{\tau_i > T\}$ given the information until time t, i.e.

$$P_\Pi(t,T) := \mathbb{E}\left[\prod_{i\in\Pi}(1 - N_i(T)) \middle| \mathcal{F}_t\right] = \mathbb{E}\left[\prod_{i\in\Pi} 1_{\tau_i > T} \middle| \mathcal{F}_t\right],$$

and $Q_\Pi(t,T) := \mathbb{E}\left[\prod_{i\in\Pi} 1_{t<\tau_i\leq T} \middle| \mathcal{F}_t\right]$. The corresponding probabilities conditional on $\bigcap_{i\in\Pi}\{\tau_i > t\}$ are denoted by

$$_{T,t}p_\Pi := \mathbb{P}\left(\tau_\Pi > T \middle| \mathcal{F}_t \wedge \{\tau_\Pi > t\}\right).$$

and $_{T,t}q_\Pi := \mathbb{P}\left(t < \tau_\Pi \leq T \middle| \mathcal{F}_t \wedge \{\tau_\Pi > t\}\right)$.

Single and joint survival probabilities will play an important role for pricing financial contracts or analyzing their risk-profiles. For example, in Chapter 4 we

calculate prices of term-life insurances, which can be considered as contingent-claims on the death time τ_i of an insured. Their values will depend on the single survival probability of the insured over the considered time period. On the other hand, *joint* survival probabilities are of particular importance when considering contracts depending on a vector of τ_i. Due to this importance, in the next two sections we provide a detailed discussion of how to derive survival probabilities within our setup.

2.2 Single Survival and Jump Probabilities

In this section, we calculate the compensators of the one-jump processes N_i and derive a general expression for the single survival probabilities.

In order to calculate single survival probabilities, we first need to compute another important quantity – the so-called *intensity* of N_i: Since $N_i(t) = 1_{\tau_i \leq t}$ is a uniformly integrable submartingale starting in 0, the famous Doob-Meyer Decomposition (see e.g. Theorem 3.15 of Jacod and Shiryaev (1987), p. 32) states that there exists a unique predictable process $A_i = (A_i(t))_{0 \leq t \leq T^*}$ (sometimes called the *dual predictable projection*) with $A_i(0) = 0$ such that

$$1_{\tau_i \leq t} - A_i(t)$$

is a uniformly integrable martingale. If $A_i(t)$ is absolutely continuous with respect to the Lebesgue measure, i.e. if A_i can be written as

$$A_i(t) = \int_0^t 1_{\tau_i > s} \lambda_i(s) ds$$

for some strictly positive, $\mathbf{F}$-adapted process λ_i, one speaks of λ_i as the *intensity process* of $1_{\tau_i \leq T}$.[6] $\lambda_i(t)\Delta$ can be interpreted as the instantaneous probability of a jump of $1_{\tau_i \leq t}$ over the next infinitesimally small time step Δ, given that the jump has not occurred yet. Under technical conditions stated in Aven (1985), the intensity of N_i can be calculated by differentiating the conditional jump probabilities:

Lemma 2.2.1 *(Aven (1985)) Let $(k_n)_{n=1}^{\infty}$ be a sequence decreasing to zero and*

$$Y_i^n(t) := \frac{1}{k_n} \mathbb{E}\left[N_i(t + k_n) - N_i(t) | \mathcal{F}_t\right].$$

Furthermore, there are non-negative and $\boldsymbol{F}$-adapted processes μ_i and y_i such that

1. *for each t and i*

$$\lim_{n \to \infty} Y_n^i(t) = \mu_i(t) \quad a.s.$$

[6] Strictly speaking, the intensity process would be $1_{\tau_i > t} \lambda_i(t)$, but in the following we refer to $\lambda_i(t)$ as the intensity process.

2. for each t and i there exists for almost all $\omega \in \Omega$ an $n_i^0 = n_i^0(t,\omega)$ such that

$$|Y_i^n(s,\omega) - \mu_i(s,\omega)| \leq y_i(s,\omega) \quad \forall s \leq t, n \geq n_i^0.$$

3. for each t and i

$$\int_0^t y_i(s)ds < \infty, \quad a.s.$$

Then,

$$N_i(t) - \int_0^t \mu_i(s)ds$$

is an $\boldsymbol{F}$-martingale.

The next proposition derives the intensity process of $1_{\tau_i \leq t}$ in our stopping times model:

Proposition 2.2.1 *Let ν_l describe the jump behavior of Λ at τ_i and let*

$$\lambda_i(s) := b_i^1(s) - \int (e^{-z_i} - 1 + z_i 1_{|z|\leq 1})\nu_{\setminus l}(s,dz)^T b_{\setminus l}^2(s)$$

be bounded. If

$$Y_i^n(s) - 1_{\tau_i > s}\lambda_i(s)$$

satisfies the conditions stated in Lemma 2.2.1, then

$$N_i(t) - \int_0^t 1_{\tau_i > s}\lambda_i(s)ds \tag{2.1}$$

is an $\boldsymbol{F}$-martingale.

Proof: For proving the martingale property of (2.1), we need an auxiliary result: Using the Itô-formula for semimartingales (cf. Bingham and Kiesel (2003), p. 212), we obtain for the bounded supermartingale $\Xi(t) := e^{-\Lambda_i(t)}$:

$$\begin{aligned}
\Xi(t) &= 1 - \int_{0+}^t e^{-\Lambda_i(u-)}d\Lambda_i(u) + \sum_{s<u\leq t}^{\Delta\Lambda_i(u)\neq 0} e^{-\Lambda_i(u-)}\left(e^{-\Delta\Lambda_i(u)} - 1\right) - e^{-\Lambda_i(u-)}\Delta\Lambda_i(u) \\
&= 1 - \int_0^t e^{-\Lambda_i(u-)}\left(b_i^1(u) - \int \left(e^{-z_i} - 1 + z_i 1_{|z|\leq 1}\right)\nu(u,dz)^T b^2(u)\right)du + M(t).
\end{aligned}$$

The last line represents the semimartingale decomposition of Ξ, where $M(t)$ denotes the martingale part. Therefore, with $t > s$ we have

$$\begin{aligned}
1 - e^{-\Lambda_i(t)+\Lambda_i(s)} &= \tilde{M}(t) + \int_s^t e^{-\Lambda_i(u)+\Lambda_i(s)}\Big(b_i^1(u) \\
&\qquad - \int \left(e^{-z_i} - 1 + z_i 1_{|z|\leq 1}\right)\nu(u,dz)^T b^2(u)\Big)du \qquad (*)
\end{aligned}$$

with $\tilde{M}(t)$ being a martingale.
Note that by definition the sets

$$\{\tau_i > t\} = \{N_i(t) = 0\} = \{\Lambda_i(t) < E_i\}$$

are equivalent and instead of considering $\{\Lambda_i(t) < E_i\}$, we may equivalently consider

$$\{\Lambda_i\left(t; n_i(t) = 0\right) < E_i\},$$

where $\Lambda_i\left(t; n_i(t) = 0\right)$ denotes that $N_i(t) = 0$ a.s. in the specification of Λ_i. In addition, in case that $n_i(t) = 0$, $\Lambda_i(t)$ and E_i are independent. Based on the exponential distribution of the E_i, by applying $(*)$ and w.l.o.g. assuming that ν_l describes the jump behavior of Λ at τ_i, we obtain

$$\begin{aligned}
\mathbb{E}[\, N_i(t+\Delta) \;-\; & N_i(t) \,|\, \mathcal{F}_t \,] = \mathbb{P}\left(t < \tau_i \leq t+\Delta | \, \mathcal{F}_t\right) = 1_{\tau_i > t}\mathbb{E}\left[\left(1 - 1_{\tau_i > t+\Delta}\right) | \, \mathcal{F}_t\right] \\
= \;& 1_{\tau_i > t}\mathbb{E}\left[1 - e^{-\Lambda_i^{\{i\}=0}(t+\Delta)+\Lambda_i^{\{i\}=0}(t)} \,\middle|\, \mathcal{F}_t\right] \\
= \;& 1_{\tau_i > t}\mathbb{E}\Bigg[\int_t^{t+\Delta} e^{-\Lambda_i^{\{i\}=0}(u)+\Lambda_i^{\{i\}=0}(t)} \Bigg(b_i^{1\{i\}=0}(u)\, du - \\
& - \int \left(e^{-z_i} - 1 + z_i 1_{|z|\leq 1}\right) \nu_{\setminus l}\left(u, dz\right)^T b_{\setminus l}^{2\{i\}=0}(u)\, du \Bigg) \,\Bigg|\, \mathcal{F}_t \Bigg],
\end{aligned}$$

where we marked a process with "$\{i\} = 0$" if in its specification $n_i(t+\Delta) = 0$ is presumed. By assumption,

$$h_i(t) := b_i^{1\{i\}=0}(t) - \int \left(e^{-z_i} - 1 + z_i 1_{|z|\leq 1}\right) \nu_{\setminus l}\left(t, dz\right)^T b_{\setminus l}^{2\{i\}=0}(t)$$

is bounded, and based on the dominated convergence theorem (DCT) we can interchange expectation and the limit, which yields

$$\begin{aligned}
\lim_{\Delta \to 0} \; & \tfrac{1}{\Delta} \; \mathbb{P}\left(t < \tau_i \leq t+\Delta | \, \mathcal{F}_t\right) \overset{DCT}{=} \\
= \;& 1_{\tau_i > t}\left(b_i^{1\{i\}=0}(t) - \int \left(e^{-z_i} - 1 + z_i 1_{|z|\leq 1}\right) \nu_{\setminus l}\left(t, dz\right)^T b_{\setminus l}^{2\{i\}=0}(t)\right) \\
= \;& 1_{\tau_i > t}\left(b_i^{1}(t) - \int \left(e^{-z_i} - 1 + z_i 1_{|z|\leq 1}\right) \nu_{\setminus l}\left(t, dz\right)^T b_{\setminus l}^{2}(t)\right),
\end{aligned}$$

where we also applied the fact that λ_i is considered on $\{\tau_i > t\}$. Thus, $n_i(t) = 0$ can be omitted in the specification of b^1 and b^2. Since all conditions of Lemma 2.2.1 are fulfilled, the claim follows.

□

In the following, we always presume that the technical conditions of Lemma 2.2.1 are satisfied. As the proposition shows, the compensator of $1_{\tau_i \leq t}$ is then given by $\int_0^t 1_{\tau_i > s} \lambda_i(s) ds$ with intensity process

$$\lambda_i(s) := b_i^1(s) - \int (e^{-z_i} - 1 + z_i 1_{|z| \leq 1}) \nu_{\setminus l}(s, dz)^T b_{\setminus l}^2(s).$$

This result leads us to the following two conclusions: First, the compensator of $1_{\tau_i \leq t}$ is continuous. Therefore (see Dellacherie and Meyer (1980)), we deal with a collection τ of totally inaccessible stopping times and work indeed in a true reduced-form setup.[7] Second, from a single object perspective, i.e. if we were only interested in modeling one default time τ_i, the jump part of Λ_i would be dispensable because we could directly specify a model for Λ_i without jump component and with a positive process $b_i^{1*}(s) := \lambda_i(s)$ instead of b_i^1. Such a model would entail exactly the same $N_i(t)$ dynamics because the one-jump process N_i is completely characterized through its intensity. Considered only from a single object's perspective, the jump part of Λ could therefore be omitted.

However, it is important to note that this is only true as long as we merely focus on the dynamics of one single N_i; from a portfolio perspective, the jump part of the jump-trigger process Λ plays a crucial role in modeling the dependence structure between the τ_is as we show in the next section.

Based on the intensity of N_i, we are now able to derive an explicit formula for the single survival probabilities. The next result has already been derived by Collin-Dufresne et al. (2004) for a setup which is comparable to ours in case that only one single N_i is considered:

Proposition 2.2.2 *For any $\mathcal{F}_T$-measurable random variable Υ satisfying*

$$\mathbb{E}\left[|\Upsilon|\right] \; < \; \infty,$$

it holds true that

$$\mathbb{E}\left[1_{\tau_i > T} \Upsilon \,|\, \mathcal{F}_t\right] = 1_{\tau_i > t} \mathbb{E}^i \left[\Upsilon e^{-\int_t^T \lambda_i(s) ds} \,\middle|\, \mathcal{F}_t \right]$$

where $\mathbb{E}^i[\cdot]$ denotes expectation with respect to an absolutely continuous measure $\mathbb{P}^i$ which is characterized by the density process

$$\Sigma_i(t) := \left. \frac{d\mathbb{P}^i}{d\mathbb{P}} \right|_{\mathcal{F}_t} = 1_{\tau_i > t \wedge T} \, e^{\int_0^{t \wedge T} \lambda_i(s) ds}.$$

[7] τ_i is called totally inaccessible if $\mathbb{P}(\omega : T(\omega) = \tau_i(\omega) < \infty) = 0$ for every predictable stopping time T (cf. Protter (2005), p. 104). Broadly speaking, this means that τ_i comes as a "surprise".

Thereby, the filtration $\boldsymbol{F}$ *is assumed to be the augmentation of the original filtration by the null sets of* $\mathbb{P}^i$*. As a special case, for* $P_i(t,T)$ *defined as in Definition 2.1.3 we obtain*

$$P_i(t,T) = 1_{\tau_i>t}\mathbb{E}^i\left[e^{-\Lambda_i(T)+\Lambda_i(t)}\middle|\,\mathcal{F}_t\right].$$

Proof: Although we proceed similar to Collin-Dufresne et al. (2004), we provide the proof in detail since our setup is slightly different. First, we prove that $\Sigma_i(t) := 1_{\tau_i>t\wedge T}\exp\left(\int_0^{t\wedge T}\lambda_i(s)ds\right)$ is an $\mathbf{F}$-martingale starting in 1. Since the martingale property on $[T,T^*]$ is self-evident, we restrict ourselves to $[0,T]$. Using the Itô-formula for semimartingales, we obtain

$$\begin{aligned} d1_{\tau_i>t}e^{\int_0^t\lambda_i(s)ds} &= 1_{\tau_i>t-}de^{\int_0^t\lambda_i(s)ds} + e^{\int_0^{t-}\lambda_i(s)ds}d1_{\tau_i>t} + 0 \\ &= 1_{\tau_i>t-}e^{\int_0^{t-}\lambda_i(s)ds}\lambda_i(t)dt + e^{\int_0^{t-}\lambda_i(s)ds}d(1-1_{\tau_i\le t}) \\ &= 1_{\tau_i>t-}e^{\int_0^{t-}\lambda_i(s)ds}\left(\lambda_i(t)dt - d1_{\tau_i\le t}\right). \end{aligned}$$

While the first equality follows from the fact that the quadratic covariation is 0, the second and third equality follow by simple rearrangements. In total, since $1_{\tau_i>t-}\left(\lambda_i(t)dt - d1_{\tau_i\le t}\right)$ is a uniformly integrable martingale and λ_i is bounded (recall that the technical assumptions stated in Proposition 2.2.1 are presumed throughout the text), $1_{\tau_i>t-}e^{\int_0^{t-}\lambda_i(s)ds}\left(\lambda_i(t)dt - d1_{\tau_i\le t}\right)$ is a uniformly integrable martingale.
Being a martingale, the process

$$\mathbb{E}\left[1_{\tau_i>t^*\wedge T}e^{\int_0^{t^*\wedge T}\lambda_i(s)ds}\middle|\,\mathcal{F}_t\right] = \Sigma_i(t) = 1_{\tau_i>t\wedge T}\,e^{\int_0^{t\wedge T}\lambda_i(s)ds}$$

with $t < t^* \le T$ is the unique density process corresponding to the absolutely continuous change of measure (cf. Theorem 3.4, Jacod and Shiryaev (1987), p. 166)

$$\frac{d\mathbb{P}^i}{d\mathbb{P}} = 1_{\tau_i>T}\,e^{\int_0^T\lambda_i(s)ds},$$

and it finally follows that

$$\begin{aligned} \mathbb{E}\left[1_{\tau_i>T}\Upsilon\middle|\,\mathcal{F}_t\right] &= 1_{\tau_i>t}e^{\int_0^t\lambda_i(s)ds}\mathbb{E}^i\left[\Upsilon e^{-\int_0^T\lambda_i(s)ds}\middle|\,\mathcal{F}_t\right] \\ &= \mathbb{E}^i\left[e^{-\int_t^T\lambda_i(s)ds}\Upsilon\middle|\,\mathcal{F}_t\right]. \end{aligned}$$

As a special case, we have

$$P_i(t,T) = 1_{\tau_i>t}\mathbb{E}^i\left[e^{-\int_t^T\lambda_i(s)ds}\middle|\,\mathcal{F}_t\right] = 1_{\tau_i>t}\mathbb{E}^i\left[e^{-\Lambda_i(T)+\Lambda_i(t)}\middle|\,\mathcal{F}_t\right],$$

since both expressions are equal because λ_i is specified through the characteristics

of Λ_i, and Λ_i cannot jump under $\mathbb{P}^i$ at τ_i, i.e. the jump part related to ν_l can be set to 0 (This can also be formally shown by applying a suitable version of the Girsanov Theorem, e.g. Theorem III.41 of Protter (2005), but is omitted here).

□

The two most important implications of Proposition 2.2.2 are the following: First, survival probabilities over the time period $[t, T]$ are given as simple expectations of an exponential of the increments of Λ_i. Second, this expectation has to be calculated with respect to a modified measure $\mathbb{P}^i$, under which N_i never jumps prior to time T $\mathbb{P}^i$-a.s because

$$\mathbb{P}^i\left(\{\tau_i \leq T\}\right) = \mathbb{E}^i\left[1_{\tau_i \leq T}\right] = \mathbb{E}\left[1_{\tau_i > T} e^{\int_0^T \lambda_i(s)ds} 1_{\tau_i \leq T}\right] = 0.$$

This means that the sets $\{\tau_i \leq T\}$ become null sets under $\mathbb{P}^i$, which is rather intuitive: As already indicated in the proof of 2.2.1, in order to compute the survival probability we need to calculate the probability that $\{\Lambda_i(T) < E_i\}$. In this computation, we have to presume $N_i(T) = 0$ for Λ_i, because by definition $\{\omega : N_i(T, \omega) = 0\} = \{\omega : \Lambda_i(T, \omega) < E_i(\omega)\}$.

We already mentioned that this intuitive yet very important result originates from Collin-Dufresne et al. (2004) (Theorem 1). Collin-Dufresne et al. (2004) were the first who showed that expectations like $\mathbb{E}\left[1_{\tau_i > T}\Upsilon \middle| \mathcal{F}_t\right]$ for some random variable Υ and some $\mathbf{F}$-stopping time τ_i can still be calculated as expectations of the form $1_{\tau_i > t} V(t)$ with $V(t) := \mathbb{E}\left[Y e^{-\int_t^T \lambda_i(s)ds} \middle| \mathcal{F}_t\right]$, even if $\Delta V_{\tau_i} \neq 0$, i.e. if V is not continuous at τ_i. However, the expectation has to be computed with respect to some measure that excludes a jump prior to T. In the literature, one usually refers to the situation $\Delta V_{\tau_i} \neq 0$ as the case where the "no jump-condition" is violated. Also, Collin-Dufresne et al. (2004) derive a formula for the expectation $\mathbb{E}\left[1_{0 < \tau_i \leq T} R_{\tau_i}\right]$ with R some predictable process. We do not state the corresponding formula for our setting since, in this thesis, we primarily focus on modeling the coordinates of τ and possible dependencies between them and we aim not at modeling what "happens" at τ_i.[8]

In our model, the stopping times are defined as the first time when $\Lambda_i(t) \geq E_i$, whereas in Collin-Dufresne et al. (2004) the stopping times are directly characterized via their intensity. The equivalence between both approaches from a single object's perspective is documented through the equation

$$1_{\tau_i > t}\mathbb{E}^i\left[e^{-\Lambda_i(T)+\Lambda_i(t)} \middle| \mathcal{F}_t\right] = 1_{\tau_i > t}\mathbb{E}^i\left[e^{-\int_t^T \lambda_i(s)ds} \middle| \mathcal{F}_t\right].$$

This result is no surprise. Remember that instead of calculating the intensity of τ_i based on our definition of τ_i, we could have directly started by specifying τ_i through

[8] Although it is possible to derive such a formula within our setup.

its intensity similar to Collin-Dufresne et al. (2004). Nonetheless, it is important to note that there remains a substantial difference between our and their setup: Although providing examples with more than one stopping time, contrarily to this thesis a general discussion of the multivariate case is missing in Collin-Dufresne et al. (2004). The examples for the multivariate case presented there do not exhaust the full flexibility of our approach. For example, simultaneous jumps of the N_i are not considered.

2.3 Joint Survival Probabilities

This section shows how joint survival probabilities can be calculated in our setup. We further derive a representation of N in terms of orthogonal point processes, i.e. point processes that never jump together. As previously indicated, for pricing financial contracts that depend on a vector of τ_i or analyzing their risk-profiles joint survival probabilities play an important role. To arrive at a formula for joint survival probabilities, we take – similar to the preceding section – the route via the intensity.

Proposition 2.3.1 *Let ν_l and ν_k describe the jump behavior of Λ at τ_i and τ_j, and let*

$$h_{ij}(t) := \sum_{k=i,j} b_k^1(t) - \int \left(e^{-z_i - z_j} - 1 + \sum_{k=i,j} z_k 1_{|z|\leq 1} \right) \nu_{\setminus\{l,k\}}(t,dz)^T \, b^2_{\setminus\{l,k\}}(t)$$

be bounded. If

$$\frac{1}{k_n}\mathbb{E}\left[1_{\tau_{(\{i,j\})}>t} - 1_{\tau_{(\{i,j\})}>t+k_n} \,\middle|\, \mathcal{F}_t \right] - 1_{\tau_{(\{i,j\})}>t} h_{ij}(t)$$

satisfies the conditions stated in Lemma 2.2.1 with k_n a sequence decreasing to zero, then

$$1 - 1_{\tau_{(\{i,j\})}>t} - \int_0^t 1_{\tau_{(\{i,j\})}>s} h_{ij}(s)ds$$

is an $\boldsymbol{F}$-martingale. Furthermore, we have that

$$1_{\tau_i=\tau_j\leq t} - \int_0^t 1_{\tau_{(\{i,j\})}>s} \lambda_{ij}(s)ds,$$

where

$$\begin{aligned} \lambda_{ij}(t) \;:=\; & \int \left(e^{-z_i - z_j} - 1 + \sum_{k=i,j} z_k 1_{|z|\leq 1} \right) \nu_{\setminus\{l,k\}}(t,dz)^T \, b^2_{\setminus\{l,k\}}(t) \\ & - \sum_{k=i,j} \int \left(e^{-z_k} - 1 + z_k 1_{|z|\leq 1} \right) \nu_{\setminus\{l,k\}}(t,dz)^T \, b^2_{\setminus\{l,k\}}(t), \end{aligned}$$

is an $\boldsymbol{F}$-martingale, too.

Proof: The proof utilizes similar ideas like the one of Proposition 2.2.1. By applying the Itô-formula for semimartingales, we arrive at the following representation of $1 - e^{-\sum_{k=i,j}\Lambda_k(t)+\Lambda_k(s)}$ for $t > s$:

$$\begin{aligned} 1 \quad - \quad e^{-\sum_{k=i,j}(\Lambda_k(t)-\Lambda_k(s))} &= \tilde{M}(t) + \int_s^t e^{-\sum_{k=i,j}(\Lambda_k(u)-\Lambda_k(s))} \left(\sum_{k=i,j} b_k^1(u) \right. \\ &\left. - \int \left(e^{-z_i-z_j} - 1 + \sum_{k=i,j} z_k 1_{|z|\leq 1} \right) \nu(u,dz)^T b^2(u) \right) du \end{aligned}$$

with $\tilde{M}(t)$ a martingale. We use this to calculate first the intensity of $1 - 1_{\tau_{(\{i,j\})}>t}$ and finally the intensity of $1_{\tau_i=\tau_j\leq t}$.
Note that by definition the events

$$\{\tau_i > t, \tau_j > t\} = \{N_i(t) = 0, N_j(t) = 0\} = \{\Lambda_i(t) < E_i, \Lambda_j(t) < E_i\}$$

are equivalent and instead of considering $\{\Lambda_i(t) < E_i, \Lambda_j(t) < E_j\}$ we may equivalently consider

$$\{\Lambda_i\left(t; n_{(\{i,j\})}(t) = 0\right) < E_i, \Lambda_i\left(t; n_{(\{i,j\})}(t) = 0\right) < E_i\},$$

where we use the same notation as in Proposition 2.2.1. Since $\Lambda_{(\{i,j\})}(t)$, E_i and E_j are independent on $\{N_{(\{i,j\})}(t) = 0\}$, we therefore obtain (E_i and E_j are independently, exponentially-distributed):

$$\begin{aligned} & \mathbb{E}[(1 - 1_{\tau_{(\{i,j\})}>t+\Delta}) - (1 - 1_{\tau_{(\{i,j\})}>t})|\mathcal{F}_t] = 1_{\tau_{(\{i,j\})}>t}\mathbb{E}\left[\left(1 - 1_{\tau_{(\{i,j\})}>t+\Delta}\right)\middle|\mathcal{F}_t\right] \\ = \quad & 1_{\tau_i>t}\mathbb{E}\left[1 - e^{-\sum_{k=i,j}\left(\Lambda_k^{\{i,j\}=0}(t+\Delta)-\Lambda_k^{\{i,j\}=0}(t)\right)}\middle|\mathcal{F}_t\right] \\ = \quad & 1_{\tau_i>t}\mathbb{E}\left[\int_t^{t+\Delta} e^{-\sum_{k=i,j}\left(\Lambda_k^{\{i,j\}=0}(u)-\Lambda_k^{\{i,j\}=0}(t)\right)} \left(\sum_{k=i,j} b_k^{1\{i,j\}=0}(u) - \right.\right. \\ & \left.\left. - \int \left(e^{-z_i-z_j} - 1 + \sum_{k=i,j} z_k 1_{|z|\leq 1} \right) \nu_{\setminus\{l,k\}}(u,dz)^T b_{\setminus\{l,k\}}^{2\{i,j\}=0}(u) \right) du|\mathcal{F}_t\right] \end{aligned}$$

where we marked a process with "$\{i,j\} = 0$" if in its specification $N_{\{i,j\}}(t+\Delta) = 0$ is assumed. By interchanging expectation and the limit based on the dominated convergence theorem (DCT), we obtain that

$$1 - 1_{\tau_{(\{i,j\})}>t} - \int_0^t 1_{\tau_{(\{i,j\})}>u} h_{ij}(u)du$$

is an $\mathbf{F}$-martingale (since the remaining conditions of Lemma 2.2.1 are satisfied by assumption) with h_{ij} defined as above. Note further that

$$\begin{aligned} d1_{\tau_i\leq t}1_{\tau_j\leq t} &= 1_{\tau_i\leq t-}d1_{\tau_j\leq t} + 1_{\tau_j\leq t-}d1_{\tau_i\leq t} + d\left[1_{\tau_i\leq t}, 1_{\tau_j\leq t}\right](t) \\ &= d1_{\tau_j\leq t} + d1_{\tau_i\leq t} - d\left(1 - 1_{\tau_{(\{i,j\})}>t}\right), \end{aligned}$$

where the first equality is due to the Itô-formula and the last line is $1_{\tau_i \leq t} 1_{\tau_j \leq t} = 1_{\tau_i \leq t} + 1_{\tau_j \leq t} - \left(1 - 1_{\tau_{(\{i,j\})} > t}\right)$ written using differential notation. Thus,

$$\begin{aligned}
d\left[1_{\tau_i \leq t}, 1_{\tau_j \leq t}\right](t) &= d1_{\tau_j \leq t} + d1_{\tau_i \leq t} - d\left(1 - 1_{\tau_{(\{i,j\})} > t}\right) - 1_{\tau_i \leq t-} d1_{\tau_j \leq t} - 1_{\tau_j \leq t-} d1_{\tau_i \leq t} \\
&= 1_{\tau_i > t-} d1_{\tau_j \leq t} + 1_{\tau_j > t-} d1_{\tau_i \leq t} - d\left(1 - 1_{\tau_{(\{i,j\})} > t}\right) \\
&\overset{(*)}{=} 1_{\tau_{(\{i,j\})} > t} \Bigg(\int \left(e^{-z_i - z_j} - 1 + \sum_{k=i,j} z_k 1_{|z| \leq 1} \right) \nu_{\setminus\{l,k\}}(t, dz)^T \, b^2_{\setminus\{l,k\}}(t) \\
&\quad - \sum_{k=i,j} \int \left(e^{-z_k} - 1 + z_k 1_{|z| \leq 1} \right) \nu_{\setminus\{l,k\}}(t, dz)^T \, b^2_{\setminus\{l,k\}}(t) \Bigg) dt + dM(t) \\
&= 1_{\tau_{(\{i,j\})} > t} \lambda_{ij}(t) dt + dM(t)
\end{aligned}$$

with λ_{ij} defined as above and where $(*)$ follows from the intensity of $1 - 1_{\tau_{(\{i,j\})} > t}$ and the intensities of $1_{\tau_i \leq t}$ ($1_{\tau_j \leq t}$) on $\{\tau_j > t-\}$ ($\{\tau_i > t-\}$) and M denotes some martingale. This finally yields the claim.

□

Remark 2.3.1 *From Proposition 2.3.1, it follows that $\lambda_{ij}(t) \equiv 0$ if Λ_i and Λ_j never jump together. In this case, simultaneous jumps of N_i and N_j cannot occur. Recently, various authors have introduced the notion of Lévy copula to characterize the dependence structure between jump processes in a dynamic way, see Chapter 5 of Cont and Tankov (2004) for a comprehensive introduction to the field of Lévy copulas. In terms of Lévy copulas, the case where two processes never jump together corresponds to the independence Lévy copula.*

Note that the case where both jump parts are 0 is included as a special case in Remark 2.3.1. At the end of Section 2.2, we have already demonstrated that when considering only one *single* stopping time τ_i the jump part of the jump-trigger process Λ_i is dispensable since we could specify a model with identical dynamics but without jumps. However, this does not hold true for *several* stopping times: Simultaneous jumps of the coordinates of N can only occur if the coordinates of Λ jump together.[9] Depending on the respective field of application, such simultaneous jumps of the N_i might be necessary to generate "enough" dependence between the N_i. For example, Longstaff and Rajan (2008) document the importance of such simultaneous jumps for explaining the price dynamics of credit portfolio derivatives.

Proposition 2.3.1 gives rise to an interesting interpretation of our stopping times model: We could calculate intensities λ_Π for all $\Pi \subseteq \{1, \ldots, I\}$ corresponding to

[9] Consequently, jumps of Λ are always dispensable if they are specified to occur independently.

the jump processes $N_\Pi := 1_{\forall\, i\neq j\in\Pi:\, \tau_i=\tau_j\leq t}$, and then recursively define the following quantities:[10]

$$\begin{aligned}
\lambda^*_{\{1,2,\ldots,I\}}(t) &:= \lambda_{\{1,2,\ldots,I\}}(t) \\
\lambda^*_{\{1,2,\ldots,I\}\setminus i}(t) &:= \lambda_{\{1,2,\ldots,I\}\setminus i}(t) - \lambda^*_{\{1,2,\ldots,I\}}(t) \\
&\vdots \\
\lambda^*_\Pi(t) &:= \lambda_\Pi - \sum_{\Pi'\subseteq\{1,\ldots,I\}:\,\Pi\subset\Pi'} \lambda^*_{\Pi'}(t) \\
&\vdots \\
\lambda^*_I(t) &:= \lambda_I - \sum_{\Pi'\subseteq\{1,\ldots,I\}:\, I\in\Pi', |\Pi'|\geq 2} \lambda^*_{\Pi'}(t)
\end{aligned} \tag{2.2}$$

The λ^*_Π represent the intensities of the one-jump processes that jump only at the point in time where all N_i, $i \in \Pi$, jump simultaneously and the remaining N_j, $j \notin \Pi$, do not jump. In the following, we denote these processes by N^*_Π, that is

$$N^*_\Pi(t) := 1_{\forall i\neq j\in\Pi,\, l\notin\Pi:\, \tau_i=\tau_j\leq t,\, \tau_i\neq\tau_l},$$

and we have that

$$N^*_\Pi(t) - \int_0^t 1_{\tau_{(\Pi)}>s}\lambda^*_\Pi(s)ds$$

is an $\mathbf{F}$-martingale. Then, each N_i can be expressed in terms of the N^*_Π as follows:

$$N_i(t) = \sum_{\Pi\subseteq\{1,\ldots,I\}} 1_{i\in\Pi} N^*_\Pi(t).$$

By definition, the one-jump processes N^*_Π are orthogonal, i.e. never jump together. The above model formulation provides a link between our setup and so-called *common Poisson shock models*, see Lindskog and McNeil (2003) for a study of this model class in an insurance and credit risk context. In these models, one directly specifies constant intensities λ^*_Π for the N^*_Π to arrive at a model of the N_i. Therefore, provided that the λ^*_Π are constant our setup also possesses an interpretation as a common Poisson shock model. We will come back to this observation in Sections 2.4 and 2.6. For the time being, we illustrate our considerations with the following example:

Example 2.3.1 Intensity Gamma *cont. (Joshi and Stacey (2006)). For ease of exposition let us consider the model introduced in Example 2.1.1 for the case $I = 2$, $p = 1$ and $a_n = a_{1n} = a_{2n}$ for each n. Based on Proposition 2.3.1, we obtain for*

[10] Presuming that the necessary, previously discussed technical conditions are fulfilled.

the intensities λ_{12}^*, λ_1^* *and* λ_2^* *with* $t \in [t_{n-1}, t_n]$*:*

$$\begin{aligned}
\lambda_{12}^*(t) &= \lambda_{12}(t) = \int \left(e^{-2\varsigma} - 1\right) \frac{ce^{-\frac{d}{a_n}\varsigma}}{\varsigma} d\varsigma - \sum_{k=1,2} \int \left(e^{-\varsigma} - 1\right) \frac{ce^{-\frac{d}{a_n}\varsigma}}{\varsigma} d\varsigma \\
&= \log\left(1 + 2\frac{a_n}{d}\right)^{-c} - 2\log\left(1 + \frac{a_n}{d}\right)^{-c} = c\log\left(\frac{\left(1 + \frac{a_n}{d}\right)^2}{\left(1 + 2\frac{a_n}{d}\right)}\right) \quad \textit{and} \\
\lambda_k^*(t) &= \lambda_1(t) - \lambda_{12}^*(t) = c\log\left(1 + \frac{a_n}{d}\right) - c\log\left(\frac{\left(1 + \frac{a_n}{d}\right)^2}{\left(1 + 2\frac{a_n}{d}\right)}\right) \\
&= c\log\left(\frac{\left(1 + 2\frac{a_n}{d}\right)}{\left(1 + \frac{a_n}{d}\right)}\right) \quad \textit{for } k = \{1, 2\}.
\end{aligned}$$

This shows that the Intensity Gamma model is in fact a common Poisson shock model with constant intensities over each time interval $[t_{n-1}, t_n]$*. The whole dependence stems from the occurrence of simultaneous jumps. At this point, it is worth mentioning that it would be easy to show that in fact every model where* Λ *is an additive process, i.e. a time-inhomogeneous "Lévy" process, can be represented as a common Poisson shock model with time-dependent intensities.*

Based on Proposition 2.3.1, we are now able to state a formula for the joint survival probability of two objects:

Proposition 2.3.2 *The joint survival probability* $P_{ij}(t,T)$ *of two portfolio objects* i *and* j *is given as*

$$P_{ij}(t,T) = 1_{\tau_{(\{i,j\})} > t} \mathbb{E}^{ij}\left[e^{-\sum\limits_{k=i,j} (\Lambda_k(T) - \Lambda_k(t))} \middle| \mathcal{F}_t \right],$$

where $\mathbb{E}^{ij}[\cdot]$ *denotes expectation with respect to an absolutely continuous measure* $\mathbb{P}^{ij}$ *which is characterized by the density process*

$$\Sigma_{ij}(t) = \left.\frac{d\mathbb{P}^{ij}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = 1_{\tau_{(\{i,j\})} > t \wedge T}\, e^{\int_0^{t\wedge T} h_{ij}(s)ds},$$

and where the filtration $\boldsymbol{F}$ *is assumed to be the augmentation of the original filtration by the null sets of* $\mathbb{P}^{ij}$*. For the joint jump probability, we obtain*

$$\mathbb{E}\left[1_{t<\tau_i,\tau_j\leq T} \middle| \mathcal{F}_t\right] = 1_{\tau_i > t} 1_{\tau_j > t} \left(P_{ij}(t,T) + 1 - P_i(t,T) - P_j(t,T)\right).$$

Proof: The martingale property of Σ_{ij} follows analogously to the martingale property of Σ_i in the proof of Proposition 2.2.2. Therefore, under $\mathbb{P}^{ij}$ $\{\tau_i \leq T\} \vee \{\tau_j \leq T\}$

becomes a null set, and we have that

$$\begin{aligned} P_{ij}(t,T) &= \mathbb{E}\left[1_{\tau_{(\{i,j\})}>T}\middle|\mathcal{F}_t\right] \\ &= 1_{\tau_{(\{i,j\})}>t}\mathbb{E}^{ij}\left[e^{-\int_t^T h_{ij}(s)ds}\middle|\mathcal{F}_t\right] \\ &= 1_{\tau_{(\{i,j\})}>t}\mathbb{E}^{ij}\left[e^{-\sum\limits_{k=i,j}(\Lambda_k(T)-\Lambda_k(t))}\middle|\mathcal{F}_t\right]. \end{aligned}$$

For the second part, we simply use that

$$1_{t<\tau_i,\tau_j\leq T} = 1_{t<\tau_i,t<\tau_j}\left(1_{\tau_i>T,\tau_j>T}+1-1_{\tau_i>T}-1_{\tau_j>T}\right).$$

By taking expectations, we obtain the claim:

$$\mathbb{E}\left[1_{t<\tau_i\leq T,t<\tau_j\leq T}\middle|\mathcal{F}_t\right] = 1_{t<\tau_i,t<\tau_j}\left(P_{ij}(t,T)+1-P_i(t,T)-P_j(t,T)\right).$$

□

It is now straight-forward to consider an arbitrary subset $\Pi\subseteq\{1,\ldots,I\}$ instead of $\{i,j\}$ in Proposition 2.3.2 and derive the corresponding joint survival probability for this set of objects. It is given as

$$P_\Pi(t,T) = 1_{\tau_{(\Pi)}>t}\,\mathbb{E}^\Pi\left[e^{-\sum\limits_{k\in\Pi}(\Lambda_k(T)-\Lambda_k(t))}\middle|\mathcal{F}_t\right], \tag{2.3}$$

where $\mathbb{E}^\Pi[\cdot]$ denotes expectation with respect to an absolutely continuous measure $\mathbb{P}^\Pi$ which is defined by the density process

$$\Sigma_\Pi = \left.\frac{d\mathbb{P}^\Pi}{d\mathbb{P}}\right|_{\mathcal{F}_t} = 1_{\tau_{(\Pi)}>t\wedge T}\,e^{\int_0^{t\wedge T}h_\Pi(s)ds},$$

and where the filtration $\mathbf{F}$ is assumed to be the augmentation of the original filtration by the null sets of $\mathbb{P}^\Pi$. Again, this formula is fairly intuitive: When calculating the joint survival probability of the objects belonging to Π, we may naturally presume that $N_i(T)=0$ for each $i\in\Pi$ because otherwise the objects would not have survived. Due to the independence of $E_{(\Pi)}$ and $\Lambda_{(\Pi)}$ in case that $N_{(\Pi)}(T)=0$ and due to the exponential distribution of $E_{(\Pi)}$, we obtain $\exp\left(-\sum\limits_{k\in\Pi}(\Lambda_k(T)-\Lambda_k(t))\right)$ in equation (2.3).

2.4 Construction of the Stopping Times

While in the previous sections we have defined a vector of stopping times and demonstrated how important quantities can be computed based on this definition, in this section we provide a construction of the considered setup, i.e. we show that

the setup is well-defined. The methodology follows Norros (1986), Shaked and Shanthikumar (1987) and Yu (2007). More precisely, we extend the construction given in Yu (2007), which is itself based on the *total hazard construction* of Norros (1986) and Shaked and Shanthikumar (1987), to the situation where simultaneous jumps of the N_i are possible. The construction represents a natural simulation algorithm for our model.

Before stating a construction algorithm, let us recall a result from the previous section. There, we showed that our model can be formulated by means of orthogonal point processes N^*_Π with intensities $1_{\tau_{(\Pi)}>s}\lambda^*_\Pi(s)$ that jump only if exactly all N_i, $i \in \Pi$, jump at the same time. The jump times of these processes are denoted by τ^*_Π in the following, i.e.

$$\tau^*_\Pi := \inf\{t : \ N^*_\Pi > 0\}.$$

Having written our setup in terms of orthogonal one-jump point processes N^*_Π, we have set the stage to apply the construction algorithm stated in Yu (2007).

To do so, we first need to introduce some notation: Following Yu (2007), we denote by $I_n = \{k_1, \ldots, k_n\}$ the identities of the n processes which have already jumped at time t; k_m identifies the process $N^*_{k_m}$ as the process which has jumped as number m. $T_n = \{t_1, \ldots, t_n\}$ denotes the set of jump times associated with the jumps of $(N_{k_1}, \ldots, N_{k_n})$ and the intensity $\lambda^*_\Pi(s) = \lambda^*_\Pi(s|I_n, T_n, X^1)$ is given as a deterministic function (defined on the interval $(t_n, t_{n+1}]$) of time, jump history and the realization of the background process X^1. The total hazard accumulated at time $t \in [t_n, t_{n+1}]$ by Π is denoted by

$$\mathcal{V}_\Pi(t|I_n, T_n, X^1) := \sum_{m=1}^{n} \mathcal{U}_\Pi(t_m - t_{m-1}|I_{m-1}, T_{m-1}, X^1) + \mathcal{U}_\Pi(t - t_n|I_n, T_n, X^1)$$

where $t_0 := 0$ and

$$\mathcal{U}_\Pi(s|I_m, T_m, X^1) = \int_{t_m}^{t_m+s} 1_{\tau_{(\Pi)}>u}\lambda^*_\Pi(u|I_m, T_m, X^1)du.$$

Furthermore, we define $\mathcal{U}^{-1}_\Pi(c|I_m, T_m, X^1) = T^*$, if $\mathcal{U}_\Pi(T^*|I_m, T_m, X^1) < c$. Based on these definitions, we are now able to state the following construction algorithm:

Algorithm 2.4.1

1. *Draw a sequence* $E = (E_1, \dots, E_I, E_{12}, \dots, E_{\{1,\dots,I\}})^T$ *of i.i.d. Exp(1)-distributed random variables.*

2. *Draw a sample path of the background process* X^1 *independent of* E.

3. *(a) Obtain* k_1 *as*

$$k_1 = \arg \min_{\Pi \subseteq \{1,\dots,I\}} \mathcal{U}_\Pi^{-1}(E_\Pi),$$

and set

$$\hat{\tau}_{k_1}^* = \mathcal{U}_{k_1}^{-1}(E_{k_1}).$$

Proceed with 3 (c).

(b) Let $(\hat{\tau}_{k_1}, \dots, \hat{\tau}_{k_{m-1}})$ *with* $m \geq 2$ *and* $T_{m-1} = \{t_1, \dots, t_{m-1}\}$ *and* $I_{m-1} = \{k_1, \dots, k_{m-1}\}$ *be given. Then,*

$$k_m := \arg \min_{\Pi \notin I_{m-1}} \mathcal{U}_\Pi^{-1} \left(E_\Pi - \mathcal{V}_\Pi(t_{m-1}|I_{m-1}, T_{m-1})| \, I_{m-1}, T_{m-1} \right)$$

and

$$\hat{\tau}_{k_m}^* := t_{m-1} + \mathcal{U}_{k_m}^{-1} \left(E_{k_m} - \mathcal{V}_{k_m}(t_{m-1}|I_{m-1}, T_{m-1})| \, I_{m-1}, T_{m-1} \right).$$

Proceed with 3 (c).

(c) If $\sum_{n=1}^m |k_n| = I$ *or* $\hat{\tau}_{k_m}^* = T^*$ *stop, otherwise increase* m *by 1 and proceed with 3 (b).*

As pointed out by Yu (2007), one can now specify a new probability space $\left(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbf{F}} = \left(\hat{\mathcal{F}}_t \right)_{0 \leq t \leq T^*}, \hat{\mathbb{P}} \right)$ where $\hat{\mathbb{P}}$ is the probability measure generated by the law of the background process X^1 and the law of $\hat{\tau}^*$ conditional on each path of X^1; $\hat{\mathbf{F}}$ is the filtration generated by the background process X^1 and the indicator processes associated with $\hat{\tau}^*$. Then, τ^* and $\hat{\tau}^*$ are equal in law conditional on each path of X^1 and the indicator processes associated with $\hat{\tau}^*$ have $\hat{\mathbf{F}}$-intensities $1_{\tau_{(\Pi)} > s} \lambda_\Pi^*(s)$. It is important to note that Algorithm 2.4.1 also represents a simulation algorithm for our stopping times model.

Furthermore, we would like to mention a slight difference between the assumptions of Yu (2007) and our setup. Since we consider in contrast to his work a final time horizon T^*, the N_i do not necessarily jump within that finite time period $[0, T^*)$. The situation where not all coordinates N_i have jumped before T^* corresponds to $\sum_{n=1}^m |k_n| < I$ and $\hat{\tau}_{k_m}^* = T^*$ in step 3 (c) of Algorithm 2.4.1.

2.5 Model-Implied Dependence Structure

After having derived formulas for single and joint survival probabilities, we now examine the dependence structure between the coordinates of N: By definition, any dependence between survival and non-survival of the portfolio objects stems from dependencies between the coordinates of the jump-trigger process Λ.[11] The joint dynamics of $(\Lambda_1, \ldots, \Lambda_I)$ determine the dynamics of the joint survival and jump probabilities.

Our main goal in this section is to analyze the dependence between the coordinates of τ over a fixed time horizon $[t, T]$ and discuss how several specifications of the jump-trigger process Λ influence this dependence structure.[12] In particular, we deal with the following questions:

- *Can we characterize the dependence structure implied by this framework?*
- *Can the dependence structure implied by our approach be compared to dependence structures generated by other approaches in the literature?*

Those two questions are of great importance when one intends to analyze differences and similarities between our approach and other approaches in the literature. For the characterization of the dependence structure, the notion of *conditional independence* is important. It is introduced in the next subsection.

2.5.1 Conditional Independence and Contagion

In the credit risk literature, there is an ongoing debate whether corporate defaults are *conditionally independent* or whether *contagion* is needed in order to explain the observed default behavior. In this thesis, we investigate this question empirically in Sections 3.1 and 3.4. Loosely speaking, contagion corresponds to the situation where the dynamics of the single objects' current default probabilities depend on past defaults of other objects and conditional independence where they do not.[13]

The rigorous mathematical definition of conditional independence, which we consider, is the following:

[11] Note that the exponentially-distributed random variables E_i, on which the definition of the τ_i is based, are i.i.d.

[12] The clustering of the jumps of N in time, i.e. an analysis of the dependence structure of N from a dynamic point of view, is examined in the next section.

[13] This is a topic which is of lower relevance in an insurance context, since in most situations it is reasonable to assume that past deaths do not influence future mortality rates, see also Subsection 2.8.2.

Definition 2.5.1 *We say that the vector $N_{(\Pi)}$ is conditionally independent on $[0,T^*]$ with respect to a subfiltration $\boldsymbol{G} \subseteq \boldsymbol{F}$* if and only if *for all $s \in [0,T^*]^I$:*

$$\mathbb{P}\left(\tau_{(\Pi)} > s_{(\Pi)} \middle| \mathcal{G}_{T^*}\right) = \prod_{i\in\Pi} \mathbb{P}\left(\tau_i > s_i \middle| \mathcal{G}_{T^*}\right).$$

This definition of conditional independence is similar to the one that can be found in Bielecki and Rutkowski (2002), p. 268. A trivial example of a filtration with respect to which $N_{(\Pi)}$ is conditionally independent is the general model filtration $\mathbf{F}$.

The interesting cases are given by true subfiltrations of $\mathbf{F}$ to which none of the coordinates of $N_{(\Pi)}$ is adapted to. These are also the setups that are usually called conditional independence models. Typically, in such a model one has common factors evolving independently of $N_{(\Pi)}$ and conditional on these factors the coordinates of $N_{(\Pi)}$ are independent. This modeling approach originates from the seminal paper of Lando (1998) and is popular in practice since it guarantees a high degree of tractability for a large range of model specifications. Note that we will tackle the issue of model tractability in Section 2.7.

The next proposition considers a particularly interesting case of conditional independence and derives the corresponding survival probabilities:

Proposition 2.5.1 *If the σ-fields $\bigvee_{i\in\Pi}\sigma(E_i)$ and $\mathcal{F}_{T^*}^{\Lambda_{(\Pi)}}$ are independent,*

1. *$N_{(\Pi)}$ is conditionally independent on $[0,T^*]$ with respect to $\boldsymbol{F}^{\Lambda_{(\Pi)}}$, and*
2. *single and joint survival probabilities are given as*

$$P_{\Pi'}(t,T) = 1_{\tau_{(\Pi')}>t}\mathbb{E}\left[\prod_{i\in\Pi'} e^{-\Lambda_i(T)+\Lambda_i(t)} \middle| \mathcal{F}_t^{N_{\setminus\Pi}} \vee \mathcal{F}_t^{\Lambda} \vee \mathcal{F}_t^{X^1}\right]$$

for any subset $\Pi' \subseteq \Pi$.

Proof: Note first that for $s \in [0,T^*]^I$ and for any sub-σ-field $\mathcal{G}$ of $\mathcal{F}$ satisfying $\mathcal{F}_{T^*}^{\Lambda_{(\Pi)}} \subseteq \mathcal{G}$ we have

$$\begin{aligned}\mathbb{P}\left(\tau_{(\Pi)} > s_{(\Pi)} \middle| \mathcal{G}\right) &= \mathbb{P}\left(\bigcap_{i\in\Pi}\{\Lambda_i(s_i) \le E_i\} \middle| \mathcal{G}\right) = \prod_{i\in\Pi} e^{-\Lambda_i(s_i)} \quad (*)\\ &= \prod_{i\in\Pi}\mathbb{P}\left(\tau_i > s_i \middle| \mathcal{G}\right),\end{aligned}$$

because the E_i are mutually independent, $\mathcal{F}_{T^*}^{\Lambda_{(\Pi)}}$ and $\bigvee_{i\in\Pi}\sigma(E_i)$ are independent and $\Lambda_{(\Pi)}(s)$ is $\mathcal{G}$-measurable. For $\mathcal{G} = \mathcal{F}_{T^*}^{\Lambda_{(\Pi)}}$, this yields the conditional independence of $N_{(\Pi)}$ with respect to $\mathcal{F}_{T^*}^{\Lambda_{(\Pi)}}$.

For the second part of the claim, we use that for an arbitrary sub-σ-field $\mathcal{K}$ of $\mathcal{F}$ and for $s > t$ the following can be shown (cf. Lemma 9.1.2 of Bielecki and Rutkowski (2002), p. 271):

$$\mathbb{P}\left(\tau_{(\Pi')} > s \middle| \mathcal{K} \vee \mathcal{F}_t^{N_{(\Pi')}}\right) = 1_{\tau_{(\Pi')}>t} \frac{\mathbb{P}\left(\tau_{(\Pi')} > s \middle| \mathcal{K}\right)}{\mathbb{P}\left(\tau_{(\Pi')} > t \middle| \mathcal{K}\right)} \quad (**).$$

Combining $(*)$ and $(**)$ we get for any subset $\Pi' \subseteq \Pi$ with $\mathcal{K} = \mathcal{F}_t^{\Lambda} \vee \mathcal{F}_t^{N_{\setminus\Pi'}} \vee \mathcal{F}_t^{X^1}$

$$\begin{aligned}
P_{\Pi'}(t,T) &= \mathbb{E}\left[1_{\tau_{(\Pi')}>T} \middle| \mathcal{K} \vee \mathcal{F}_t^{N_{(\Pi')}}\right] \\
&\overset{(**)}{=} 1_{\tau_{(\Pi')}>t} \frac{\mathbb{E}\left[\mathbb{E}\left[1_{\tau_{(\Pi')}>T} \middle| \mathcal{F}_{T^*}^{\Lambda_{(\Pi')}} \vee \mathcal{F}_t^{\Lambda_{\setminus\Pi'}} \vee \mathcal{F}_t^{N_{\setminus\Pi'}} \vee \mathcal{F}_t^{X^1}\right] \middle| \mathcal{K}\right]}{\mathbb{E}\left[1_{\tau_{(\Pi')}>t} \middle| \mathcal{K}\right]} \\
&\overset{(*)}{=} 1_{\tau_{(\Pi')}>t} \mathbb{E}\left[\prod_{i\in\Pi'} e^{-\Lambda_i(T)+\Lambda_i(t)} \middle| \mathcal{K}\right] \\
&= 1_{\tau_{(\Pi')}>t} \mathbb{E}\left[\prod_{i\in\Pi'} e^{-\Lambda_i(T)+\Lambda_i(t)} \middle| \mathcal{F}_t^{N_{\setminus\Pi}} \vee \mathcal{F}_t^{\Lambda} \vee \mathcal{F}_t^{X^1}\right],
\end{aligned}$$

where we also used the law of iterated expectations. The last equality follows from the fact that $\bigvee_{i\in\Pi, i\notin\Pi'} \sigma(E_i)$ and $\mathcal{F}_{T^*}^{\Lambda_{(\Pi')}}$ are independent (for a detailed argument see Lando (1998)). Alternatively, the survival probabilities could be derived by applying Propositions 2.2.2 and 2.3.2 and using the Girsanov theorem afterwards in order to show that the dynamics of $\Lambda_{(\Pi)}$ are the same under $\mathbb{P}$ and $\mathbb{P}^{\Pi}$.

□

If the conditions of Proposition 2.5.1 are fulfilled, conditional on the history of the jump-trigger process $\Lambda_{(\Pi)}$ the coordinates of $N_{(\Pi)}$ are independent. Note that this does not imply that jumps of $N_{\setminus\Pi}$ cannot influence the survival probabilities of $N_{(\Pi)}$ or vice versa but it implies that the jumps of the N_i with $i \in \Pi$ cannot affect each other. It is worth mentioning that such a model includes as a special case a setup investigated in Jarrow and Yu (2001). In the so-called *Primary-Secondary* credit portfolio framework considered there, defaults of *primary* firms ($j \notin \Pi$) can affect the default probabilities of *secondary* firms ($i \in \Pi$) but not vice versa.

The next definition introduces the notion of *contagion* and distinguishes between different types of contagion.

Definition 2.5.2 Contagion *is* not *present among the coordinates of* $N_{(\Pi)}$ if and only if *the* σ*-fields* $\bigvee_{i\in\Pi} \sigma(E_i)$ *and* $\mathcal{F}_{T^*}^{\Lambda_{(\Pi)}}$ *are independent. Furthermore, considering*

a fixed time horizon $[t,T]$*, we say that* positive (negative) contagion *is present among the coordinates of* $N_{(\Pi)}$ *in* $[t,T]$ *if for all* $i \in \Pi$ *and* $a_i > 0$ *we always have*

$$\mathbb{E}^i\left[\exp\left(-a_i\left(\Lambda_i(T)-\Lambda_i(t)\right)\right)\middle|\,\mathcal{F}_t\right] \overset{(\geq)}{\leq} \mathbb{E}^{\Pi}\left[\exp\left(-a_i\left(\Lambda_i(T)-\Lambda_i(t)\right)\right)\middle|\,\mathcal{F}_t\right],$$

and if in addition there exists at least one $j \in \Pi$ *for which*

$$\mathbb{E}^j\left[\exp\left(-a_j\left(\Lambda_j(T)-\Lambda_j(t)\right)\right)\middle|\,\mathcal{F}_t\right] \overset{(>)}{<} \mathbb{E}^{\Pi}\left[\exp\left(a_j\left(-\Lambda_j(T)-\Lambda_j(t)\right)\right)\middle|\,\mathcal{F}_t\right]$$

holds true.

Definition 2.5.2 determines contagion not to be present among a subset Π of objects if the independence assumption stated in Proposition 2.5.1 is satisfied. As already mentioned, this assumption guarantees that the survival probabilities of all objects $i \in \Pi$ do not depend on jumps of $N_{(\Pi)}$. Following the literature, we will henceforth exchangeably use the terms conditional independence setup, setup without contagion or doubly stochastic setup to denote such a model. Contrarily, a model where this assumption is violated will be called a setup with contagion or a contagion model.

Also, Definition 2.5.2 distinguishes between different types of contagion. Positive (negative) contagion is said to be present among a subset of portfolio objects if the possibility of other defaults $j \in \Pi, j \neq i$ within this subset reduces (increases) the survival chances of each object $i \in \Pi$ over a fixed time period, since the probability of large realizations of $\Lambda_i(T) - \Lambda_i(t)$ is increased (decreased). This means that other defaults may have (and have at least in case of one object) a negative (positive) impact on the survival probability of the surviving firms. Furthermore, the direction of this impact is for each object the same: Either its survival probability is affected in a negative (positive) way or it is not affected at all. When characterizing the model-implied dependence structure in the next subsection, we will get back to the situation when positive (negative) contagion is present among the objects and discuss the implications for the dependence structure.

Table 2.1 summarizes general classifications of the model-implied dependence structure. As pointed out in the preceding section, simultaneous jumps of Λ_i and Λ_j, $i \neq j$, imply a positive probability for simultaneous jumps of N_i and N_j, and the independence of the σ-fields $\bigvee_{i\in\Pi}\sigma(E_i)$ and $\mathcal{F}_{T^*}^{\Lambda_{(\Pi)}}$ implies a setup where contagion is ruled out among the objects belonging to Π.

	$\bigvee_{i\in\mathbf{\Pi}} \boldsymbol{\sigma}(\boldsymbol{E_i})$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{T_*}}^{\boldsymbol{\Lambda_{(\Pi)}}}$ **ind.**	$\bigvee_{i\in\mathbf{\Pi}} \boldsymbol{\sigma}(\boldsymbol{E_i})$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{T_*}}^{\boldsymbol{\Lambda_{(\Pi)}}}$ **not ind.**
No simult. jumps of $\boldsymbol{\Lambda_{(\Pi)}}$	No contagion, No simult. jumps of $N_{(\Pi)}$	Contagion, No simult. jumps of $N_{(\Pi)}$
Simult. jumps of $\boldsymbol{\Lambda_{(\Pi)}}$	No contagion, Simult. jumps of $N_{(\Pi)}$	Contagion, Simult. jumps of $N_{(\Pi)}$

Table 2.1: General classifications of the model-implied dependence structure between the coordinates of $N_{(\Pi)}$ with respect to contagion and simultaneous jumps.

2.5.2 A Characterization Result for the Dependence Structure

Next, we describe the dependence structure between the components of τ in more detail. We establish a link between our time-continuous setup and the static concept of *copulas* – a popular tool in the context of dependence modeling. In recent years, copulas (or equivalently copula functions) have enjoyed increasing popularity in the finance community for basically one reason: They represent a simple, ad-hoc method to introduce dependence between random variables; given the margins of a random vector, one only needs to specify an arbitrary copula function to obtain a multivariate distribution function for the vector. In an application, for example, this allows for calibrating the margins and the dependence between the single random variables in two separate steps significantly simplifying the estimation procedure.

The idea behind the characterization of the model-implied dependence structure that follows now, is simple and possibly best described by an example. Consider a time-continuous model consisting of two correlated Brownian motions. We could now "switch" from the time-continuous to a static setting by analyzing only the distribution of the Brownian motions at the end of some fixed time interval as implied by the time-continuous model. This distribution is a two-dimensional Gaussian distribution and dependence between both Brownian motions is described by a Gaussian copula. In the following, we conduct exactly the same analysis for our – by far more complicated – time-continuous model of multiple stopping times.

The motivation supporting our investigation is twofold: First, the established link allows us to use results from the well-developed copula theory for a further analysis of the dependence structure of concrete model specifications. Second, the link

makes our model better comparable with other models in the literature, especially to the static copula models, see e.g. Li (2000) for a credit portfolio model based on the Gaussian copula and Burtschell et al. (2007) for a comparative analysis of different copula functions applied to the pricing of credit portfolio derivatives.[14]

Important Definitions and Results from the Theory of Copula Functions

If not stated otherwise, the following standard definitions, results and their proofs can be found in Nelsen (2006), Joe (1997) or McNeil et al. (2005), where also further properties and details regarding copulas not discussed in this thesis are presented. A copula function is formally defined as follows:

Definition 2.5.3 *A function* $C : [0,1]^I \to [0,1]$ *is called a* copula *if it has the following properties:*

1. $C(1,\ldots,1,u_i,1,\ldots,1) = u_i$ *for all* $i = 1,\ldots,I$, $u_i \in [0,1]$ *and* $C(u) = 0$ *for every* $u \in [0,1]^I$ *if at least one coordinate of the vector* u *is 0;*

2. *For all* a,$b \in [0,1]^I$ *with* $a_i \leq b_i$, $1 \leq i \leq I$, *the volume of the hypercube with corners* a, b *is positive:*

$$\sum_{j_1=1}^{2}\sum_{j_2=1}^{2}\cdots\sum_{j_I=1}^{2}(-1)^{j_1+j_2+\ldots+j_I}C(u_{1_{j_1}},u_{2_{j_2}},\ldots,u_{I_{j_I}}) \geq 0$$

where $u_{i_1} = a_i$ *and* $u_{i_2} = b_i$ *for all* $i = 1,\ldots,I$.

Definition 2.5.3 implies that a copula function is a multivariate distribution function with uniform margins. The central theorem in the context of copula functions is Sklar's theorem. It establishes the link between copula function, joint distribution function and the marginal distributions of random vectors.

Theorem 2.5.1 ***(Sklar)*** *Let* $V_1, V_2, \ldots, V_I$ *be random variables with marginal distribution functions* $G_1, G_2, \ldots, G_I$ *and joint distribution function* G. *Then, there exists an* I*-dimensional copula* C *such that for all* $v \in \mathbb{R}^I$

$$G(v_1,\ldots,v_I) = C(G_1(v_1),\ldots,G_I(v_I)).$$

If $G_1, G_2, \ldots, G_I$ *are continuous, then* C *is unique. Otherwise* C *is uniquely determined on* $Ran(G_1) \times Ran(G_2) \times \cdots \times Ran(G_I)$, *where* $Ran(G_i)$ *denotes the range of* G_i *for* $i = 1,2,\ldots,I$.

[14] The copula method has also been applied in a time-continuous setting, see e.g. Schönbucher and Schubert (2001). However, while its application seems reasonable in a static setup where only one fixed time horizon is considered and where one essentially deals with random variables its application to a dynamic framework is more than questionable: For example, Rogge and Schönbucher (2003) point out that the copula method applied to a dynamic credit portfolio context can result in completely unrealistic model dynamics.

Sklar's theorem highlights the aforementioned power of the copula concept in dependence modeling: On the one hand, copulas allow to separate the dependence part of the joint distribution function, described by the copula function, from the marginal distributions. On the other hand, a joint distribution function can be constructed by transforming given margins by an arbitrary copula.

A particular class of copulas, which will be important for the characterization of the model-implied dependence structure, is given by the next definition:

Definition 2.5.4 *Let be given an I-dimensional vector $V = (V_1, \ldots, V_I)^T$ of positive random variables. Then, we call the copula function defined by*

$$\begin{aligned} C(u_1, \ldots, u_I) &= \mathbb{E}\left[\prod_{i=1}^{I} \exp\left(-V_i \varphi_i^{-1}(u_i)\right)\right] \qquad \textit{with} \\ \varphi_i(x) &= \mathbb{E}\left[\exp\left(-xV_i\right)\right] \qquad 1 \leq i \leq I \end{aligned} \tag{2.4}$$

a copula of the Marshall-Olkin *type.*

The copula introduced in Definition 2.5.4, is a special case of an even more general copula class introduced in Theorem 2.1 of Marshall and Olkin (1988), p. 835. It is important to note that the copula of the Marshall-Olkin type considered here and the so-called *Marshall Olkin copula* (see e.g. Nelsen (2006), p. 53, for a definition) represent different copula functions.

Remark 2.5.1 *A further – particularly nice – subclass of the Marshall-Olkin type copulas are the LT-Archimedean copulas, which have been studied by Rogge and Schönbucher (2003) and McNeil et al. (2005).*[15] *In case of a LT-Archimedean copula, each V_i can be decomposed as*

$$V_i = \sum_{j=1}^{p} A_{ij} U_j, \tag{2.5}$$

i.e. each V_i exhibits a certain factor structure, where $A \in \mathbb{R}_{0+}^{I \times p}$, and the factors U_j are assumed to be positive, independent random variables. In the special case $p = 1$ and $A_{i1} = A_{k1}$ for all $i \neq k$, equation (2.5) reduces to

$$C(u_1, \ldots, u_I) = \varphi\left(\sum_{i=1}^{I} \varphi^{(-1)}(u_i)\right). \tag{2.6}$$

In the literature, such a copula is called an Archimedean copula function and one refers to $\phi = \varphi^{-1}$ as its generator.[16] *It is worth mentioning that the copula functions constructed by means of equation (2.6) constitute only a subset of the ordinary*

[15] "LT" stands for Laplace transform indicating that the copulas are constructed based on Laplace transforms of positive random variables.

[16] To discern between the general and the special case we refer to them as LT-Archimedean and ordinary or simple Archimedean copulas in the following.

Archimedean copula functions since the Archimedean copula class is usually defined in a broader sense (cf. Nelsen (2006) or Joe (1997)).

As a consequence of its definition, a LT-Archimedean copula function can be written as

$$C(u_1, \ldots, u_I) = \prod_{j=1}^{p} \mathbb{E}\left[\exp\left(-U_j \sum_{i=1}^{I} A_{ij} \varphi_i^{(-1)}(u_i)\right)\right].$$

Characterization of the Model-Implied Dependence Structure

For the analysis of the model-implied dependence structure, we restrict ourselves now to an arbitrary, fixed time horizon $[t, T]$. Thereby, we focus on the following questions:

Given that we scale the Λ_is introduced in Definition 2.1.1 by positive constants a_i, what is the function that maps the different individual survival probabilities corresponding to different values $a = (a_1, \ldots, a_I)^T$ to the joint survival probability? Can we characterize the mechanism linking single and joint survival probabilities in this case?

These questions are not only interesting from an academic point of view. They become particularly important when calibrating a model to a given data set. To further illustrate this point, let us consider the following potential credit portfolio application of our stopping times model: Its calibration to prices of securities that reference the whole portfolio and to prices of securities that only reference the single objects in the portfolio. As already indicated in connection with Example 2.1.1, the most natural way of doing this would be to scale the Λ_i by positive constants a_i to match the different levels of individual prices and adjust the parameters that describe the joint dynamics of the jump-trigger processes Λ_i to fit the prices of the securities referencing the whole portfolio. In Mortensen (2006) and Feldhütter (2008), for example, such a scaling is applied in the calibration of the Duffie and Gârleanu (2001) credit portfolio model.

The scaling of the Λ_i intrinsically changes the measure $\mathbb{P}$ since now

$$\tau_i := \inf\{t : a_i \Lambda_i(t) \geq E_i\}$$

and we write $\mathbb{P}^a$ (or $\mathbb{P}^{\Pi,a}$ for $\Pi \subseteq \{1, \ldots, I\}$ if we change to one of the measures considered in the previous sections).[17] As the probability measure is changed, even single survival probabilities may depend on the whole vector a (changing the

[17]Note that model calibration is per se the procedure of finding the "right" measure. Usually, one a priori restricts the set of possible measures. Here, we restrict ourselves to the measures which correspond to a scaling of the Λ_i.

survival probability of one object can change the survival probability of another object, too) and we write ${}_{T,t}p_i(a)$ instead of ${}_{T,t}p_i$ to emphasize the dependence of the survival probabilities on a. Moreover, whenever we restrict ourselves to a subset of objects $\Pi \subseteq \{1,\ldots,I\}$, different combinations of a_i are only considered for $i \in \Pi$, while the $a_j \equiv 1$, $j \notin \Pi$, are kept fixed.

We now focus on the mechanism linking the different possible marginal survival probabilities ${}_{T,t}p_i(a)$ of a subset Π of objects to their joint survival probability, given the information up to time t and leaving the specification of Λ itself unchanged. We are interested in the following function:

Definition 2.5.5 *Consider the mapping* $P : a \mapsto {}_{T,t}p_{(\Pi)}(a)$ *with* $a \in \mathbb{R}_{0+}^{|\Pi|}$ *and denote by RAN(P) its range. We call the function* ${}_{T,t}F_\Pi : RAN(P) \to [0,1]$ *with property*

$$ {}_{T,t}p_\Pi(a) = {}_{T,t}F_\Pi\left({}_{T,t}p_{(\Pi)}(a)\right) \qquad \forall a \in \mathbb{R}_{0+}^{|\Pi|}, $$

the dependence function of $N_{(\Pi)}$.[18]

Remark 2.5.2 *We limit our attention to the dependence function conditional on survival of the objects* Π *in order to simplify notation; this is not a real restriction since the "unconditional" function* F *satisfying* $P_\Pi(t,T) = F\left(P_{(\Pi)}(t,T)\right)$ *coincides with* ${}_{T,t}F_\Pi$ *on* $\bigcap_{i\in\Pi}\{\tau_i > t\}$.

In connection with the dependence function, it is important to note the following: First, the function ${}_{T,t}F_\Pi$ should not be mistaken for the model-implied survival times copula, which is often considered in the literature (for an example see Georges et al. (2001)). The survival times copula C is the copula linking the marginal survival probabilities over different time horizons to the joint survival probability, i.e.

$$ \mathbb{P}\left(\tau > T \middle| \mathcal{F}_t \wedge \cap_{i\in\Pi}\{\tau_i > t\}\right) = C\left({}_{T_1,t}p_1, \ldots, {}_{T_I,t}p_I\right) $$

with $T = (T_1,\ldots,T_I)^T > t$. Contrarily, we consider the function which maps possible combinations of survival probabilities over the same time horizon to the joint survival probability. Nevertheless, for fixed a and $\overline{T} = T_1 = \ldots = T_I$ both functions yield the same value, i.e.

$$ {}_{\overline{T},t}p_{\{1,\ldots,I\}} = \mathbb{P}\left(\tau > \overline{T} \middle| \mathcal{F}_t \wedge \cap_{i\in\Pi}\{\tau_i > t\}\right) = C\left({}_{\overline{T},t}p_1, \ldots, {}_{\overline{T},t}p_I\right). $$

From a practical point of view, the dependence function is the more important quantity since in an application we are usually first interested in the dependence

[18] Watch the notation: ${}_{T,t}p_{(\Pi)}$ denotes the $|\Pi|$-dimensional sub-vector of $({}_{T,t}p_1, \ldots, {}_{T,t}p_I)^T$ with elements ${}_{T,t}p_i$, $i \in \Pi$ while ${}_{T,t}p_\Pi$ denotes the joint conditional survival probability of all $i \in \Pi$.

between $N_i(T)$ and $N_j(T)$ and to a lesser extent in dependencies between $N_i(T_i)$ and $N_j(T_j)$, $i \neq j$.

Second, the dependence function is not the copula function linking the probabilities of the Bernoulli-events {"survival", "non-survival"}, but this copula can be expressed in terms of the function.

Third, focusing on the survival events is not a restriction since the function ${}_{T,t}\overline{F}_{\Pi}$ with $\Pi = \{i, j\}$ for the jump events can easily be obtained from ${}_{T,t}F_{\{i,j\}}$ (higher dimensional cases can be treated similarly):

$$ {}_{T,t}\overline{F}_{\{i,j\}} \left({}_{T,t}q_i \, , \, {}_{T,t}q_j \right) = 1 - {}_{T,t}p_i - {}_{T,t}p_j + {}_{T,t}F_{\{i,j\}} \left({}_{T,t}p_{(\{i,j\})} \right). $$

For the characterization of ${}_{T,t}F_{\Pi}$, Laplace transforms of the increments of Λ will play a crucial role:

Definition 2.5.6 ${}^{\Pi',a}_{T,t}\varphi_{\Lambda_{(\Pi)}}$ *with* $\Pi' \subseteq \{1, \ldots, I\}$ *is defined as the joint Laplace transform under the measure* $\mathbb{P}^{\Pi',a}$ *of the increments of* $\Lambda_{(\Pi)}$ *over the time interval* $[t, T]$ *conditional on the information until time* t*, i.e.*

$$ {}^{\Pi',a}_{T,t}\varphi_{\Lambda_{(\Pi)}} \left(u_{(\Pi)} \right) = \mathbb{E}^{\Pi',a} \left[e^{-\sum\limits_{i \in \Pi} u_i (\Lambda_i(T) - \Lambda_i(t))} \middle| \mathcal{F}_t \right] $$

with $u \in \mathbb{R}^I_{0+}$.

With Definition 2.5.6 at hand, we have set the stage for a first, very general characterization result for the dependence function ${}_{T,t}F_{\Pi}$ of $N_{(\Pi)}$ over $[t, T]$:

Proposition 2.5.2 *The dependence function* ${}_{T,t}F_{\Pi}$ *is given as*

$$ {}_{T,t}F_{\Pi} \left({}_{T,t}p_{(\Pi)}(a) \right) = {}^{\Pi}_{T,t}\varphi_{\Lambda_{(\Pi)}} \left(u_{(\Pi)} \right), $$

where

$$ u = \left({}^{1,a}_{T,t}\varphi^{-1}_{\Lambda_1} \left({}_{T,t}p_1(a) \right), \ldots, {}^{I,a}_{T,t}\varphi^{-1}_{\Lambda_I} \left({}_{T,t}p_I(a) \right) \right). $$

Proof: From Proposition 2.2.2 it directly follows for all $i \in \{1, \ldots, I\}$ that

$$ {}_{T,t}p_i(a) = {}^{i,a}_{T,t}\varphi_{a_i \Lambda_i}(1) = {}^{i,a}_{T,t}\varphi_{\Lambda_i}(a_i) \qquad \Rightarrow \qquad {}^{i,a}_{T,t}\varphi^{-1}_{\Lambda_i} \left({}_{T,t}p_i(a) \right) = a_i, $$

since the Laplace transform is invertible and the distribution of $\Lambda_i(T) - \Lambda_i(t)$ does under $\mathbb{P}^{i,a}$ not depend on a_i. Furthermore, based on Proposition 2.3.2 we obtain

$$ {}_{T,t}p_{\Pi}(a) = {}^{\Pi}_{T,t}\varphi_{\Lambda_{(\Pi)}} \left(u_{(\Pi)} \right) = {}_{T,t}F_{\Pi} \left({}_{T,t}p_{(\Pi)}(a) \right) $$

with $u = \left({}^{1,a}_{T,t}\varphi^{-1}_{\Lambda_1}\left({}_{T,t}p_1(a)\right), \ldots, {}^{I,a}_{T,t}\varphi^{-1}_{\Lambda_I}\left({}_{T,t}p_I(a)\right)\right)^T$ as claimed.

□

Note that ${}^{\Pi}_{T,t}\varphi_{\Lambda_{(\Pi)}}$ in Proposition 2.5.2 does not depend on a anymore and "a" is dropped because under the measure $\mathbb{P}^{\Pi}$ jumps of the objects Π do not occur before T (recall that all a_j, $j \notin \Pi$ are kept fixed). a does therefore have no influence on the distribution of $\Lambda_i(T) - \Lambda_i(t)$. Proposition 2.5.2 highlights the importance of the increments of Λ for the dependence structure of N: The dependence structure is completely determined by the multivariate and univariate Laplace transforms of these increments under the different measures.

The next theorem shows that for a setup without contagion the dependence function is a copula function. More precisely, the dependence function is a copula of the Marshall-Olkin type (cf. Definition 2.5.4) in this case.

Theorem 2.5.2 *The model-implied dependence function ${}_{T,t}F_{\Pi}$ defines a copula function of the Marshall-Olkin type if among the coordinates $N_{(\Pi)}$ contagion is not present in the sense of Definition 2.5.2.*

Proof: Given conditional independence, we can use Proposition 2.5.1 to get for the single survival probabilities:

$$ {}_{T,t}p_i(a) = {}^{i,a}_{T,t}\varphi_{\Lambda_i}(a_i) = {}_{T,t}\varphi_{\Lambda_i}(a_i) = {}_{T,t}p_i(a_i). $$

This means that the Laplace transform is independent of a, and since $\Lambda_i(T) - \Lambda_i(t)$ is a positive random variable we therefore find for each $p \in [0,1]^{|\Pi|}$ an $a \in \mathbb{R}^{|\Pi|}_{0+}$ such that

$$ {}_{T,t}\varphi_{\Lambda_i}(a_i) = p_i, \qquad \text{for all} \qquad i \in \Pi, $$

by setting $a_i = {}_{T,t}\varphi^{-1}_{\Lambda_i}(p_i)$. This implies that $RAN(P) = [0,1]^{|\Pi|}$, i.e. in a conditional independence model any combination of $p \in [0,1]^{|\Pi|}$ can be reproduced by scaling the Λ_i and the dependence function is defined on $[0,1]^{|\Pi|}$.
Based on the conditional independence, we finally obtain for the dependence function ${}_{T,t}F_{\Pi}$ that

$$ \begin{aligned} {}_{T,t}p_{\Pi} &= \mathbb{E}^{\Pi}\left[\prod_{i\in\Pi} e^{-a_i(\Lambda_i(T)-\Lambda_i(t))}\middle|\mathcal{F}_t\right] \\ &= \mathbb{E}\left[\prod_{i\in\Pi} e^{-a_i(\Lambda_i(T)-\Lambda_i(t))}\middle|\mathcal{F}_t\right] \\ &= \mathbb{E}\left[\prod_{i\in\Pi} e^{-\,{}_{T,t}\varphi^{-1}_{\Lambda_i}\left({}_{T,t}p_i\right)(\Lambda_i(T)-\Lambda_i(t))}\middle|\mathcal{F}_t\right] = {}_{T,t}F_{\Pi}\left({}_{T,t}p_{(\Pi)}\right) \end{aligned} $$

where the expectation and the transforms ${}_{T,t}\varphi_{\Lambda_i}$ are calculated with respect to the same measure $\mathbb{P}$, which proves that the dependence function is a copula of the Marshall-Olkin type (cf. Definition 2.5.4).

□

In connection with Theorem 2.5.2 we can state the subsequent corollary:

Corollary 2.5.1

1. *The function* ${}_{T,t}C_\Pi : [0,1]^\Pi \to [0,1]$ *with*

$$ {}_{T,t}C_\Pi\,(u) := {}^{\Pi}_{T,t}\varphi_{\Lambda_{(\Pi)}}\left(\left({}^{\Pi}_{T,t}\varphi^{-1}_{\Lambda_1}\,(u_1), \ldots, {}^{\Pi}_{T,t}\varphi^{-1}_{\Lambda_I}\,(u_I)\right)_{(\Pi)}\right) \tag{2.7}$$

   always defines a copula function of the Marshall-Olkin type.

2. *If for all* $i \in \Pi$ $\Lambda_i(t)$ *can be written under* $\mathbb{P}^\Pi$ *as*

$$\Lambda_i(t) = \sum_{j=1}^{p} A_{ij}\Lambda^*_j(t),$$

   where the Λ^*_j *are of the general form considered in equation (2.1), i.e. each* Λ^*_j *only consists of a drift and a jump part, and if*

$$\mathbb{E}^\Pi\left[e^{-\sum\limits_{j=1}^{p} A_{ij}\left(\Lambda^*_j(T)-\Lambda^*_j(t)\right)}\,\middle|\,\mathcal{F}_t\right] = \prod_{j=1}^{p}\mathbb{E}^\Pi\left[e^{-A_{ij}\left(\Lambda^*_j(T)-\Lambda^*_j(t)\right)}\,\middle|\,\mathcal{F}_t\right]$$

   for all $i \in \Pi$ *and arbitrary* $A_{ij} \geq 0$*, then* ${}_{T,t}C_\Pi$*, defined in equation (2.7), is an LT-Archimedean copula. In the special case* $p = 1$ *and* $A_{i1} = A_{k1}$ *for* $k \neq i \in \Pi$*,* ${}_{T,t}C_\Pi$ *is an ordinary Archimedean copula function.*

Proof: The first part follows from Definition 2.5.4, and the second part follows from the definition of the LT-Archimedean and the ordinary Archimedean copula function (cf. Remark 2.5.1).

□

Theorem 2.5.2 and Corollary 2.5.1 establish an important link between our dynamic setup and the static copula approach. In the static copula approach, a particular copula function is exogenously imposed in order to link the marginal survival probabilities to the joint survival probability. We, however, specified a dynamic model for the jump times of N and studied then its implied dependence structure over a fixed time horizon. We found that in our stopping times model the dependence

mechanism which transforms single survival probabilities to joint survival probabilities is closely related to a copula function which has already been investigated by Marshall and Olkin (1988).

In the contagion case, the function linking single and joint survival probabilities is not a copula since the Laplace transforms are calculated with respect to different measures in this case. Also, in such a model possibly not every combination of single survival probabilities can be reproduced by multiplying the processes Λ_i with some positive factor. To see this, consider the following very simple two object example which is again taken from the field of credit portfolio risk modeling: Over a fixed time horizon, firm 1 survives with probability p_1, and firm 2 survives whenever 1 survives. In addition, firm 2 survives alone with some probability p_*. Then of course, firm 2 has survival probability $p_2 = p_1 + p_*$, which can never be smaller than p_1. Furthermore, the dependence function linking both survival probabilities is in this case only defined on $(p_1, p_1 + p_*)$ for all $p_1, p_* > 0$ such that $p_1 + p_* \leq 1$.

If a model does not exhibit contagion, Theorem 2.5.2 states that the single and joint survival probabilities are connected as implied by a copula function of the Marshall-Olkin type. Model calibration as described in this section corresponds to the calibration of a copula of a Marshall-Olkin type. A particular subclass of this very broad class of copula functions is given by the LT-Archimedean and by the ordinary Archimedean copulas. They are obtained by imposing a certain structure of independent factors on the Λ_i. It is worth noting that many credit portfolio models which have been proposed in the literature, are based on such a factor structure and therefore imply a LT-Archimedean dependence structure, see e.g. the models by Mortensen (2006), Joshi and Stacey (2006), Graziano and Rogers (2006), Papageorgiou and Sircar (2007) or Chapovsky et al. (2006). It further follows that any one-factor model that does not exhibit contagion – as considered in the second part of Corollary 2.5.1 – implicitly assumes an ordinary Archimedean copula dependence structure. Examples are the model by Graziano and Rogers (2006) or the one-factor versions of the other models above. We come back to this point in Section 2.8, where we will explicitly derive the copula function of the Duffie and Gârleanu (2001) model.

Therefore, Theorem 2.5.2 and Corollary 2.5.1 demonstrate that the dependence structures implied by many time-continuous models proposed in literature are comparable to the ones of static Archimedean copula models.

The next proposition compares models that exhibit positive or negative contagion to models not featuring contagion effects.

Proposition 2.5.3 *If there exists positive contagion among $N_{(\Pi)}$ in the sense of Definition 2.5.2 for all $a \geq 0$, we have the following relationship between the model-implied dependence function ${}_{T,t}F_{\Pi}$ and the copula function ${}_{T,t}C_{\Pi}$ defined by equation (2.7):*

$$ {}_{T,t}C_{\Pi} \quad \prec \quad {}_{T,t}F_{\Pi}. $$

Proof: By Definition 2.5.2, we can find an $j \in \Pi$ such that for all $a_j \geq 0$:

$$ {}_{T,t}p_j\left(a\right) = {}^{j,a}_{T,t}\varphi_{\Lambda_j}\left(a_j\right) < {}^{\Pi}_{T,t}\varphi_{\Lambda_i}\left(a_j\right) $$

and for $i \in \Pi, i \neq j$, the relation holds with a $\leq$ in place of $<$. Since the Laplace transform is strictly decreasing,

$$ {}^{\Pi}_{T,t}\varphi^{-1}_{\Lambda_j}\left({}_{T,t}p_j(a)\right) > {}^{j,a}_{T,t}\varphi^{-1}_{\Lambda_j}\left({}_{T,t}p_j(a)\right) = a_j. $$

This finally yields that

$$ \begin{aligned} {}_{T,t}C_{\Pi}\left({}_{T,t}p_{(\Pi)}\left(a\right)\right) &= \mathbb{E}^{\Pi}\left[e^{-\sum\limits_{i\in\Pi}{}^{\Pi}_{T,t}\varphi^{-1}_{\Lambda_i}\left({}_{T,t}p_i\left(a\right)\right)\left(\Lambda_i(T)-\Lambda_i(t)\right)}\middle|\mathcal{F}_t\right] \\ &\leq \mathbb{E}^{\Pi}\left[e^{-\sum\limits_{i\in\Pi}{}^{i,a}_{T,t}\varphi^{-1}_{\Lambda_i}\left({}_{T,t}p_i\left(a\right)\right)\left(\Lambda_i(T)-\Lambda_i(t)\right)}\middle|\mathcal{F}_t\right] \\ &= {}_{T,t}F_{\Pi}\left({}_{T,t}p_{(\Pi)}\left(a\right)\right) \end{aligned} $$

for all $a \geq 0$.

□

Of course, the counterpart holds true in case of negative contagion. Again, the result is fairly intuitive: The dependence between the default and survival events increases if positive contagion is present compared to a setting without contagion. On the other hand, negative contagion lowers dependence and has a diversification effect: The occurrence of jumps of $N_{(\Pi)}$ make further jumps more unlikely.

In practical applications of a model, one often has to rely on Monte-Carlo simulations for calculating integrals if analytical solutions are not available. In this case, a simulation algorithm is needed. We end our discussion of the model-implied dependence structure by providing an algorithm for the simulation of random variables which are distributed according to the copula function ${}_{T,t}C_{\Pi}$. It is inspired by an algorithm stated in Marshall and Olkin (1988):

Algorithm 2.5.1 *Follow the algorithm:*

1. *Simulate $|\Pi|$ random variables X_i, $1 \leq i \leq |\Pi|$, uniform on $[0,1]$, i.i.d.*
2. *Simulate $\Lambda_{(\Pi)}(T) - \Lambda_{(\Pi)}(t)$ under the measure $\mathbb{P}^{\Pi}$ conditional on the information until time t, expressed through the sigma-field $\mathcal{F}_t$.*

3. Define

$$U_i := {}^{\Pi}_{T,t}\varphi_i \left(\frac{1}{\Lambda_i(T) - \Lambda_i(t)} \cdot (-\ln X_i) \right) \qquad i \in \Pi.$$

Then, the joint distribution function of the U_i is given by ${}_{T,t}C_{\Pi}$.

Dependence Measures

In the following, we introduce and briefly discuss several dependence measures, which we will subsequently apply (see e.g. Section 2.8). The first measure is the linear correlation coefficient between the jump events:

Definition 2.5.7 *The jump correlation ${}_T\rho_t^{ij}$ of two objects $i \neq j$ is defined as the linear correlation coefficient between the jump events:*[19]

$${}_{T,t}\rho_{ij} := \frac{{}_{T,t}q_{ij} - {}_{T,t}q_i \cdot {}_{T,t}q_j}{\sqrt{{}_{T,t}q_i \,(1 - {}_{T,t}q_i\,)\; {}_{T,t}q_j \,(1 - {}_{T,t}q_j\,)}}$$

Note that jump and survival correlation are equivalent and depend on the dependence function introduced in Definition 2.5.5, as the following simple transformation shows:

$$\begin{aligned} {}_{T,t}\rho_{ij} &= \frac{{}_{T,t}q_{ij} - {}_{T,t}q_i \cdot {}_{T,t}q_j}{\sqrt{{}_{T,t}q_i \,(1 - {}_{T,t}q_i\,)\; {}_{T,t}q_j \,(1 - {}_{T,t}q_j\,)}} \\ &= \frac{1 - {}_{T,t}p_i - {}_{T,t}p_j + {}_{T,t}p_{ij} - (1 - {}_{T,t}p_i)(1 - {}_{T,t}p_j\,)}{\sqrt{{}_{T,t}p_i \,(1 - {}_{T,t}p_i\,)\; {}_{T,t}p_j \,(1 - {}_{T,t}p_j\,)}} \\ &= \frac{{}_{T,t}F_{ij}\,(\,{}_{T,t}p_i\,,\,{}_{T,t}p_j\,) - {}_{T,t}p_i\;{}_{T,t}p_j}{\sqrt{{}_{T,t}p_i \,(1 - {}_{T,t}p_i\,)\; {}_{T,t}p_j \,(1 - {}_{T,t}p_j\,)}}. \end{aligned}$$

Therefore, for two different models 1 and 2 it holds true that

$${}_{T,t}\rho_{ij}^1 \prec {}_{T,t}\rho_{ij}^2$$

if ${}_{T,t}F_{ij}^1 \prec {}_{T,t}F_{ij}^2$. Besides linear correlation, we will consider the following dependence measures:

[19] In the credit risk related literature, the jump correlation is usually called the *default correlation*.

Definition 2.5.8 *Given two random variables X^1, X^2 with copula C,*

1. *the* lower *and* upper tail dependence *coefficient, ζ_L and ζ_U, are defined as*

$$\begin{aligned} \zeta_L &= \lim_{u\to 0} \frac{C(u,u)}{u}, \\ \zeta_U &= \lim_{u\to 1} \frac{1-u-u+C(1-u,1-u)}{1-u}, \end{aligned}$$

   given that the limits exist.

2. Kendall's tau Ψ *is defined as*

$$\Psi = 4\int_{[0,1]^2} C(u_1,u_2)dC(u_1,u_2) - 1.$$

For a discussion of advantages and disadvantages of the different dependence measures, see e.g. Embrechts et al. (2002). Both measures introduced in Definition 2.5.8 are copula-based, i.e. can only be calculated if no contagion is present among $\{i,j\}$ and the model-implied dependence function is a copula (cf. Theorem 2.5.2). Nevertheless, at least the computation of the upper and lower tail dependence coefficient is feasible if the dependence function is defined for all ${}_{T,t}p_i = {}_{T,t}p_j$ as it is usually the case in applications.

In case of an Archimedean copula function C the tail dependence coefficient as well as Kendall's tau can conveniently be calculated as follows:

$$\begin{aligned} \Psi &= 1 + 4\int_0^1 \frac{\varphi^{-1}(u)}{\frac{\partial}{\partial u}\varphi^{-1}(u)}du, \\ \zeta_l &= \lim_{u\to 0} \frac{2\frac{\partial}{\partial u}\varphi(2\varphi^{-1}(u))}{\frac{\partial}{\partial u}\varphi(\varphi^{-1}(u))}, \\ \zeta_u &= 2 - \lim_{u\to 1} \frac{2\frac{\partial}{\partial u}\varphi(2\varphi^{-1}(u))}{\frac{\partial}{\partial u}\varphi(\varphi^{-1}(u))}. \end{aligned}$$

For a proof of the first equality see Corollary 5.1.4 of Nelsen (2006), p. 163, and to obtain the second equality apply l'Hospital's rule.

2.6 The Loss Process

The perspective which we have taken so far was the static portfolio perspective, i.e. we considered each single jump process N_i and described the joint distribution of these one-point jump processes over a fixed time period. Subsequently, we directly consider the aggregated process L, which we refer to in the following as the *loss process* and which is defined as

$$L(t) := \sum_{i=1}^{I} 1_{\tau_i \leq t} = \sum_{i=1}^{I} N_i(t).$$

By definition, L is a bounded, pure jump process which only jumps at a maximum of I different points with jump sizes between 1 and I. The first contribution of this section is the characterization of the dynamics of L. Furthermore, we provide a time-change result which will be at the bottom of an important statistical test considered in Section 3.1. As a second contribution, we propose a new measure to characterize the volatility of L and discuss the properties of this measure.

2.6.1 Formulation by Means of Orthogonal Point Processes and Time-Change

With L being a uniformly integrable submartingale, the first natural question that arises is about the compensator A of the process. Due to the linearity of the expectation operator, this compensator is given as

$$A(t) = \sum_{i=1}^{I} A_i(t) = \sum_{i=1}^{I} \int_0^t \lambda_i(s) 1_{\tau_i > s} ds, \tag{2.8}$$

i.e. $L(t) - \sum_{i=1}^{I} \int_0^t \lambda_i(s) 1_{\tau_i > s} ds$ is an $\mathbf{F}$-martingale, with λ_i as in Proposition 2.2.1. However, the compensator alone does not characterize L, because simultaneous jumps of the coordinates of N may occur and $\Delta L(t) \leq 1$ does therefore not hold true in general. Such a complete characterization of L can be obtained by calculating its characteristics.

Instead of calculating these characteristics, we will consider a different quantity, which eventually contains the same information in the case examined here. This quantity is the compensator of the quadratic variation of L. The quadratic variation process of L, $[L, L]$, is given by

$$[L, L](t) = \sum_{0 < s \leq t}^{\Delta L(s) \neq 0} (\Delta L(s))^2$$

since L is a pure jump process; it equals L if the jump size of L is always 1. $[L, L](t) = L(t)$ therefore corresponds to the situation where $[N_i, N_j](t) = 0$, for

all $i, j \in \{1, \ldots, I\}$, $i \neq j$ and $t > 0$. Generally, we can derive the following decomposition of L (cf. Section 2.3):

$$L(t) = \sum_{k=1}^{I} H_k(t) \cdot k,$$

where H_k is defined as

$$H_k(t) := \sum_{\Pi \subseteq \{1,\ldots,I\}:\, |\Pi|=k} N^*_\Pi(t).$$

Hence, H_k counts the jumps of L with size k. By construction, we have that

$$[H_k, H_l](t) = 0 \qquad \forall t \in [0, T^*],$$

that is for $k \neq l$ H_k and H_l are *orthogonal*. With $[L, L]$ being a pure jump process, we can again consider its compensator, $\langle L, L \rangle$, which can be written as:

Proposition 2.6.1 *The compensator $\langle L, L \rangle$ of the quadratic variation of L is given by*

$$\begin{aligned} \langle L, L \rangle(t) &= \sum_{k=1}^{I} k^2 \sum_{\Pi \subseteq \{1,\ldots,I\}:\, |\Pi|=k} \int_0^t 1_{\tau_{(\Pi)} > u} \lambda^*_\Pi(u) du \\ &= \sum_{k=1}^{I} k^2 \int_0^t \lambda^\perp_k(u) du, \end{aligned}$$

where

$$\lambda^\perp_k(u) := \sum_{\Pi \subseteq \{1,\ldots,I\}:\, |\Pi|=k} 1_{\tau_{(\Pi)} > u} \lambda^*_\Pi(u).$$

Note that in connection with Proposition 2.3.1 we presented a recursive scheme for the computation of the λ^*_Π.

As already mentioned, in later applications we will make use of a time-change result regarding the portfolio loss process L. The basis of this time-change is the following theorem due to Meyer (1971)):

Theorem 2.6.1 *(Meyer (1971)) If a multivariate point process $(\Phi_1, \ldots, \Phi_K)$ has a continuous compensator $(A_1, \ldots, A_K)$ such that $A_k(\infty) = \infty$ for each $k = 1, \ldots, K$, then the point processes $(\Phi^*_1, \ldots, \Phi^*_K)$ defined by*

$$(\Phi^*_1(t), \ldots, \Phi^*_K(t)) := \left(\Phi_1\left(A_1^{-1}(t)\right), \ldots, \Phi_K\left(A_K^{-1}(t)\right)\right)$$

are independent Poisson processes.

Note that a multivariate point process is a point process whose coordinates are orthogonal (see e.g. Brémaud (1981)). The next corollary is a direct consequence of Theorem 2.6.1 applied to our setup:

Corollary 2.6.1 *The inter-arrival times of the jumps of*

$$\Phi_k^*(t) := H_k\left(A_k^{-1}(t)\right), \qquad 1 \leq k \leq I,$$

with

$$A_k(s) = \sum_{\Pi \subseteq \{1,\ldots,I\}:|\Pi|=k} \int_0^s \lambda_\Pi^*(u)\, 1_{\tau_{(\Pi)}>u} du = \int_0^s \lambda_k^\perp(u) du$$

form a finite sequence of independently, Exp(1)-distributed random variables, with $\lambda_k^\perp$ defined as in Proposition 2.6.1.

Note that in Corollary 2.6.1 $A_k(\infty) = \infty$ does not hold true, but as pointed out by Brown and Nair (1988) Theorem 2.6.1 can easily be generalized to this situation. In this case, we obtain a finite sequence of exponentially distributed random variables and not an infinite series, i.e. a Poisson process.

Corollary 2.6.1 will be at the bottom of our statistical tests which we consider in Subsection 3.1.4 when we explore the question of whether a path of observed portfolio defaults is likely to have been generated by an estimated model of the default intensities.

2.6.2 Volatility Structure of the Point Process and a Measure of Default Clustering

In the preceding section, we characterized the dependence structure between the N_i over a fixed time horizon. This investigation was eventually static in the sense that we considered the question: Given the information up to time t what is the dependence structure between the random variables $N_i(T)$? While the dependence structure explored there is closely related to the distribution of $L(T)$ *at the end* of the period $[t, T]$, in the following, we rather want to characterize the paths of L *over* $[t, T]$.

To motivate the following investigation, let us consider the following simple example of two point processes N_1' and N_2' over the time period $[0, 2t]$ both having jumps of size 1: While the jumps of N_2' arrive with intensities $\lambda_2' = const.$, N_1' jumps according to a stochastic intensity λ_1': With probability p the intensity is $\lambda_1' = \lambda_2'$, and with probability $1 - p$ we have that $\lambda_1'(s) = c_1 > 0$ for all $s \in [0, t]$ and $\lambda_1'(s) = c_2 > 0$ for all $s \in [t, 2t]$. Moreover, we assume that $c_2 >> c_1$ and set $\lambda_2' = \frac{c_1+c_2}{2}$. Then,

$$\mathbb{E}\left[N_1'(2t)\right] = \mathbb{E}\left[N_2'(2t)\right] = \lambda_2' 2t = Var\left[N_1'(2t)\right] = Var\left[N_2'(2t)\right]$$

because $N_1'(2t)$ and $N_2'(2t)$ are identically distributed. Nevertheless, from a dynamic perspective N_1' intuitively represents the riskier process since with probability $1-p$ jumps of N_1' can be expected to be heavily clustered in $[t,2t]$, while the jumps of N_2' will always be relatively evenly distributed over $[0,2t]$. Imagine, for example, that N_1' and N_2' model the loss process of two different credit portfolios, 1 and 2. Then, the portfolio with a higher probability of a large number of losses within a small time interval, i.e. portfolio 1, is the riskier one from an investor's perspective. This higher risk of portfolio 1 is intimately linked to the path properties of N_1' whose paths are likely to show a more volatile behavior than the paths of N_2'.

Our example shows that in a time-continuous setup it can be insufficient to merely focus on traditional risk measures such as the expected loss or the variance of the loss process, which are based on the distribution of L at the end of a particular time period $[t,T]$. In such a setup, also the path behavior of L over $[t,T]$ should be examined because this behavior is connected to the default clustering within $[t,T]$. We propose the following measure in order to describe a volatile path behavior of the loss process L:

Definition 2.6.1 *The* expected volatility ${}_{T,t}\sigma_L$ *of the loss process L over a time period $[t,T]$ is defined by*

$$\begin{aligned}
{}_{T,t}\sigma_L &:= \mathbb{E}\left[\left(\left[\sum_{k=1}^{I} k\lambda_k^{\perp}+L,\sum_{k=1}^{I} k\lambda_k^{\perp}+L\right](T)-\left[\sum_{k=1}^{I} k\lambda_k^{\perp}+L,\sum_{k=1}^{I} k\lambda_k^{\perp}+L\right](t)\right)\middle|\mathcal{F}_t\right]\\
&= \mathbb{E}\left[\left(\left\langle\sum_{k=1}^{I} k\lambda_k^{\perp}+L,\sum_{k=1}^{I} k\lambda_k^{\perp}+L\right\rangle(T)-\left\langle\sum_{k=1}^{I} k\lambda_k^{\perp}+L,\sum_{k=1}^{I} k\lambda_k^{\perp}+L\right\rangle(t)\right)\middle|\mathcal{F}_t\right].
\end{aligned}$$

By simple transformations, we obtain

$$\begin{aligned}
{}_{T,t}\sigma_L &= \mathbb{E}\left[\left(\left\langle\sum_{k=1}^{I} k\lambda_k^{\perp},\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(T)-\left\langle\sum_{k=1}^{I} k\lambda_k^{\perp},\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(t)\right)\middle|\mathcal{F}_t\right]\\
&\quad+\mathbb{E}\left[2\cdot\left(\left\langle L,\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(T)-\left\langle L,\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(t)\right)+\left(\langle L,L\rangle(T)-\langle L,L\rangle(t)\right)\middle|\mathcal{F}_t\right]\\
&= \mathbb{E}\left[\left(\left\langle\sum_{k=1}^{I} k\lambda_k^{\perp},\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(T)-\left\langle\sum_{k=1}^{I} k\lambda_k^{\perp},\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(t)\right)\middle|\mathcal{F}_t\right]\\
&\quad+\mathbb{E}\left[2\cdot\left(\left\langle\sum_{k=1}^{I} k\cdot H_k,\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(T)-\left\langle\sum_{k=1}^{I} k\cdot H_k,\sum_{k=1}^{I} k\lambda_k^{\perp}\right\rangle(t)\right)\middle|\mathcal{F}_t\right]\\
&\quad+\mathbb{E}\left[\left(\left\langle\sum_{k=1}^{I} k\cdot H_k,\sum_{k=1}^{I} k\cdot H_k\right\rangle(T)-\left\langle\sum_{k=1}^{I} k\cdot H_k,\sum_{k=1}^{I} k\cdot H_k\right\rangle(t)\right)\middle|\mathcal{F}_t\right]. \quad (2.9)
\end{aligned}$$

${}_{T,t}\sigma_L$ measures the expected variation of the jumps of L. This jump clustering depends on the loss process as well as on its intensity and possible dependencies between both. Each of the three terms of equation (2.9) carries different information on the volatility of L: The first represents the variation of L's intensity, the second accounts for dependencies between the loss process and its compensator and the third contains the contributions of the loss process L itself to ${}_{T,t}\sigma_L$. It is worth noting that in case of a conditional independence setup without simultaneous jumps it will be sufficient to consider only the first term, since the second will be 0 and the third does not carry any information on the dependence structure in this case. Also, with ${}_{t+\Delta,t}\sigma_L$ being an $\mathbf{F}$-adapted process reflecting the expected jump clustering in $[t, t+\Delta]$, it is worthwhile to study its dynamics. If ${}_{t+\Delta,t}\sigma_L$ does not vary a lot, the clustering of the jump times stays relatively stable over time, too. In particular, in case of a homogeneous model, i.e. $\lambda_k^{\perp}(t) \equiv \lambda_k^{\perp}$ for all $t \geq 0$ or a time-inhomogeneous but deterministic model, where the $\lambda_k^{\perp}(t)$ are only deterministic functions of time, ${}_{t+\Delta,t}\sigma_L$ will be also deterministic.[20]

The third term of (2.9) accounts for simultaneous jumps of the objects: Larger jumps of L enter with a higher weight, since

$$[L, L](t) = \sum_{k=1}^{I} k^2 H_k \geq \sum_{k=1}^{I} k \cdot H_k = L(t).$$

This results in the nice property that two different models whose loss process compensator is identical but where model 1 does allow for jump sizes larger than 1 and model 2 does not, we always have

$${}_{T,t}\sigma_{L_1} \geq {}_{T,t}\sigma_{L_2},$$

showing that expected volatility as a measure of default clustering reacts as desired in this case.

At this point, we would like to stress that we do not suggest to use the expected volatility on its own in order to compare the riskiness of different models. Rather, we propose to augment the set of measures, which is used so far to compare two different models, by the expected volatility. In addition, when examining a particular model it is useful to study the dynamics of ${}_{T,t}\sigma_L$. If expected volatility fluctuates a lot, the default clustering in time and thus the dependence structure will vary a lot, too.

We conclude this discussion by returning to our initial, motivating example for which we find that

$${}_{2t,0}\sigma_{N_2'} = \lambda_2' 2t < \lambda_2' 2t + (1-p)(c_2 - c_1) = {}_{2t,0}\sigma_{N_1'}.$$

[20] For instance, in the Intensity Gamma model, which has been studied in Examples 2.1.1 and 2.3.1, ${}_{t+\Delta,t}\sigma_L$ evolves deterministically.

Conditional Independence vs. Contagion

We now apply ${}_{T,t}\sigma_L$ introduced in Definition 2.6.1 in order to analyze differences between the jump behavior of a loss process L whose default intensities do not depend on past defaults, i.e. a conditional independence model, and a process whose jumps exhibit a feedback property, i.e. a model with contagion.

For simplicity of exposition, we assume that

$$L(t) = \sum_{k=1}^{I} k \cdot H_k(t)$$

as above but where the compensator intensities of the H_ks are now – different to their initial definition in Proposition 2.6.1 – given as

$$\lambda_k^{\perp}(t) := \sum_{\Pi \subseteq \{1,\dots,I\}:|\Pi|=k} \lambda_{\Pi}^{*}(t)\,.$$

This assumption is solely made to keep notation simple by avoiding the downward jumps of the intensities after a default, which can be observed in a conditional independence as well as in a contagion setup; the assumption implies that the number of jumps of L is not bounded by I anymore. Technically (cf. Definition 7 of Brémaud (1981)), we have to presume in this case

$$\int_0^t \lambda_{\Pi}^{*}(s)ds < \infty \qquad \forall t \in [0, T^*] \tag{2.10}$$

such that explosions are avoided.

Being the sum of processes that depend on the background process X^1 and on the contagion process X^2, the $\lambda_k^{\perp}$ are again functions of X^1 and X^2. For simplicity, we assume that

$$\lambda_k^{\perp}(t) = \mu_k^1(X^1(t)) + \sum_{0<s\le t}^{\Delta L(s)\neq 0} c\,\Delta L(s)\, e^{-\kappa(t-s)}, \tag{2.11}$$

where c and κ are positive constants, X^1 as always evolves independently of L – that is the σ-fields $\mathcal{F}_{T^*}^{X^1}$ and $\mathcal{F}_t^N$ are independent conditional on $\mathcal{F}_t^{X^1}$ – and μ_k^1 is a continuous function defined on the state space of X^1.[21] Then, the compensator of $\left[\sum_{k=1}^I k\lambda_k^{\perp}, \sum_{k=1}^I k\lambda_k^{\perp}\right]$ is given by

[21] For more general specifications of the intensities, a qualitatively identical result holds, but the notation becomes substantially more involved without providing further insights.

$$\begin{aligned}\left\langle \sum_{k=1}^{I} k\lambda_k^{\perp}, \sum_{k=1}^{I} k\lambda_k^{\perp} \right\rangle (t) &= \left\langle \sum_{k=1}^{I} k\mu_k^1, \sum_{k=1}^{I} k\mu_k^1 \right\rangle (t) \\ &\quad + \left(c \sum_{k=1}^{I} k \right)^2 \left\langle \sum_{0<s\leq \bullet}^{\Delta L(s)\neq 0} \Delta L(s)\,, \sum_{0<s\leq \bullet}^{\Delta L(s)\neq 0} \Delta L(s) \right\rangle (t) \\ &= \left\langle \sum_{k=1}^{I} k\mu_k^1, \sum_{k=1}^{I} k\mu_k^1 \right\rangle (t) + \left(\sum_{k=1}^{I} k \right)^2 \int_0^t \sum_{n=1}^{I} (cn)^2 \lambda_n^{\perp}(s) ds \\ &> \left\langle \sum_{k=1}^{I} k\mu_k^1, \sum_{k=1}^{I} k\mu_k^1 \right\rangle (t).\end{aligned}$$

For the contagion component of the expected volatility we obtain

$$\begin{aligned}\left\langle L, \sum_{k=1}^{I} k\lambda_k^{\perp} \right\rangle (t) &= \left(c \sum_{k=1}^{I} k \right) \cdot \left\langle \sum_{0<s\leq \bullet}^{\Delta L(s)\neq 0} \Delta L(s), \sum_{0<s\leq \bullet}^{\Delta L(s)\neq 0} \Delta L(s) \right\rangle (t) \\ &= c\,\frac{I(I+1)}{2} \int_0^t \sum_{k=1}^{I} k^2 \lambda_k^{\perp}(s) ds \;>\; 0\end{aligned}$$

and the compensator of $[L, L]$ has the form

$$\langle L, L \rangle (t) \;=\; \int_0^t \sum_{k=1}^{I} k^2 \lambda_k^{\perp}(s) ds \;>\; \int_0^t \sum_{k=1}^{I} k^2 \mu_k^1(s) ds.$$

This shows that all three components of the expected volatility are increased compared to a model that does not exhibit joint jumps of the intensity and the loss process L, i.e. a model where $\lambda_k^{\perp}(t) = \mu_k^1(X^1(t))$ for each k. Adding positive contagion effects to a conditional independence specification therefore increases the expected volatility, which is not surprising since an increase can be expected whenever an additional risk source is added to a specification.

But what is the difference between adding the loss process and adding an independent jump process as an extra risk driver? To further explore this question, let us consider an intensity specification of the form

$$\lambda_k^{\perp(2)}(t) = \mu_k^1(X^1(t)) + \sum_{0<s\leq t}^{\Delta J(s)\neq 0} c\,\Delta J(s)\, e^{-\kappa(t-s)}, \tag{2.12}$$

where we assume $[J, L] = 0$, $J = \sum_{k=1}^{I} k \cdot H_k^{(2)}$ and the $H_k^{(2)}$ to be orthogonal point processes with intensities $\lambda_k^{\perp(2)}$. This model represents a conditional independence model because the $\lambda_k^{\perp(2)}$ evolve independently of L. Furthermore, the random variables $\langle L, L \rangle (t)$ and $\left\langle \sum_{k=1}^{I} k\lambda_k^{\perp}, \sum_{k=1}^{I} k\lambda_k^{\perp} \right\rangle (t)$ have in both models given by

equations (2.11) and (2.12) the same distribution (and hence the same expectation). The only difference is that in case of the latter $\left\langle \sum_{k=1}^{I} k\lambda_k^{\perp(2)}, L^{(2)} \right\rangle (t) = 0$, which still implies that

$$ {}_{T,t}\sigma_{L^{(2)}} \; < \; {}_{T,t}\sigma_{L} $$

for a conditional independence model (${}_{T,t}\sigma_{L^{(2)}}$) and a model with contagion (${}_{T,t}\sigma_{L}$), even if the intensity processes $\lambda_k^{\perp}$ show in both models the same marginal dynamics. When characterizing the jump clustering of the loss process, it is therefore important not only to consider the volatility of its intensity process but also to take possible dependencies between both into account.

We conclude this section by emphasizing that the results presented in this subsection do not suggest that conditional independence models are not able to provide enough flexibility for modeling a certain pattern of jump clustering. Rather, they show that expected volatility as a measure of jump clustering reacts in a reasonable way: Adding observed defaults as an additional risk source increases the measure. Furthermore, we showed that it is not sufficient to consider merely the dynamics of the intensity in order to assess the jump clustering, but also possible dependencies between intensity and loss process have to be taken into account.

The question of whether conditional independence models provide enough modeling pliability or if contagion models are coercively necessary in order to gain a sufficient degree of flexibility for modeling the dynamics of credit portfolios will be further investigated in Sections 3.1 and 3.4.

2.7 Analytically Tractable Model Specifications

So far, we have characterized the model-implied dependence structure from a static as well as from a dynamic point of view. We have also obtained general formulas for important quantities such as survival probabilities. In this section, we investigate the question for which type of model specifications these general formulas can analytically or at least semi-analytically be solved. This represents an important issue, since analytical tractability of a model greatly facilitates its implementation and closed-form solutions always have the advantage that parameter sensitivities can easily be computed.

The first contribution of this section is the calculation of an important, general transform. This very general result could also be useful in a totally different context than considered in this thesis and is based on the works by Duffie et al. (2000) and Duffie et al. (2003). In a second step, we then show how the result can be applied to compute the quantities which are of particular interest in this thesis.

2.7.1 A General Formula for Determining the Characteristic Function of a Process

It is well-known that the class of so-called exponential-affine processes has many appealing features. In particular, their characteristic function and further important transforms are known in closed-form up to the solution of generalized Riccati equations (see Duffie et al. (2000) and Duffie et al. (2003)). Not surprisingly, one therefore often relies on an affine model when it comes to a concrete model specification; for examples of affine credit portfolio models see Hurd and Kuznetsov (2007) and Chen and Filipović (2007), and for affine model examples in an insurance context see Biffis (2005), Dahl (2004) or Schrager (2006).

In this subsection, we show that one can eventually obtain a comparable degree of analytical tractability in a more flexible framework than given by the exponential-affine model class. Before stating the main result of this subsection, we need to make some assumptions:

Assumption 2.7.1 *Let $X = (Y^1, Y^2, Z)$ be a $(d_1' + d_2' + 1)$-dimensional Markov process (in the following $d' := d_1' + d_2' + 1$) on a filtered probability space $\left(\Omega', \mathcal{F}', \boldsymbol{F} = (\mathcal{F}_t')_{0\le t\le T^*}, \mathbb{P}'\right)$ with state space $\mathbb{R}^{d_1'} \times \mathbb{R}_+^{d_2'} \times D$, where D denotes the finite state space $\{e_1, \ldots, e_r\}$ of Z with $e_m = (1_{1=m}, 1_{2=m}, \ldots, 1_{r=m})$ for all $m \in \{1, \ldots, r\}$. We assume that X evolves according to the following stochastic differential equation (SDE):*

$$dX(t) \;=\; \beta(t, X(t))dt + \sigma(t, X(t))dW(t) + dJ(t)$$

where W denotes a d'-dimensional Brownian motion and J a jump process whose jumps are characterized through $\vartheta(t, X(t-); dt, dx)$. We assume that with $x = (y^1, y^2, z) \in \mathbb{R}^{d_1'} \times \mathbb{R}_+^{d_2'} \times D$

$$\begin{aligned}
\beta(t,x) &= z\beta^1(t) + \beta^2(t)y \\
\sigma(t,x)\sigma(t,x)^T &= z\gamma^1(t) + \gamma^2(t)y^2 \\
\vartheta(t,x;dt,d\varsigma) &= z\nu^1(t,d\varsigma) + \nu^2(t,d\varsigma_1,\ldots,d\varsigma_{d'-1})y^2dt.
\end{aligned}$$

Furthermore, $\beta^1, \beta^2, \gamma^1, \gamma^2, \nu^1$ and ν^2 are r-dimensional and time-dependent such that for each $t \ge 0$ and $m \in \{1, \ldots, r\}$ the following holds:

$$\begin{aligned}
\beta_m^1(t) &\in \mathbb{R}^{d_1'} \times \mathbb{R}_{0+}^{d_2'} \times \{0\} \\
\beta^2(t) &\in \mathbb{R}^{d'\times(d'-1)} \text{ with restrictions}: \\
&\qquad \beta_{kl}^2(t) \equiv 0 \quad d_1' < k \le d'-1,\ 1 \le l \le d_1' \\
&\qquad \beta_{kl}^2(t) > 0 \quad d_1' < k \ne l \le d'-1 \\
&\qquad \beta_{d'l}^2(t) \equiv 0 \quad 1 \le l \le d'-1
\end{aligned}$$

$$
\begin{aligned}
\gamma_m^1(t) \;&\in \mathbb{R}^{d'\times d'} \text{ a symmetric, positive semidefinite matrix} \\
&\quad \text{with restriction} \\
&\quad \gamma_{mkl}^1(t) \equiv 0 \quad d_1' < k, l \\
\gamma^2(t) \;&\in \mathbb{R}^{d'\times d'\times d_2'}, \text{ where for each } l \in \{1,\dots,d_2'\} \ \left(\gamma_{k_1k_2l}^2\right)_{1\le k_1,k_2\le d'} \\
&\quad \text{a symmetric, positive semidefinite matrix whose entries} \\
&\quad \text{satisfy for } d_1' < k_1, k_2 < d' \\
&\quad \gamma_{k_1k_2l}^2(t) \begin{cases} \geq 0 & , \quad k_1 = k_2 = (l + d_1') \\ = 0 & , \quad \text{else} \end{cases}, \\
&\quad \text{and } \gamma_{k_1k_2l}^2 = 0 \text{ for } k_1 = d' \text{ or } k_2 = d'. \\
\nu_m^1(t) \quad &\text{is a Lévy measure satisfying for each } t \geq 0 \\
&\quad \nu_m^1(t, \mathbb{R}^{d_1'} \times \mathbb{R}_+^{d_2'} \times D') < \infty \\
\nu_l^2(t) \quad &\text{is a Lévy measure satisfying for each } l \in \{1,\dots,d_2'\} \text{ and } t \geq 0 \\
&\quad \nu_l^2(t, \mathbb{R}^{d_1'} \times \mathbb{R}_+^{d_2'}) < \infty
\end{aligned}
$$

with $D' := \{e_n - e_m : n \neq m\}$.

Remark 2.7.1 *The imposed parameter restrictions of Assumption 2.7.1 regarding* (Y^1, Y^2) *are those from Duffie et al. (2003) who state them for the exponential affine case. For a detailed discussion we therefore refer to their work. Note that for each fixed state* m *of* Z, (Y^1, Y^2) *belongs to the class of exponential affine processes studied in Duffie et al. (2003). The introduced restrictions ensure, for example, that* C *is a symmetric non-negative matrix and that* Y^2 *is a non-negative process. Our restrictions regarding the Lévy measures entail that the jumps of* X *show finite activity. It is possible to relax this assumption and consider jumps with infinite activity. In this case, additional restrictions stated in Duffie et al. (2003) have to be taken into account.*

An Ornstein-Uhlenbeck process Y^1 and a CIR process Y^2 (see Section 3.2), which are both driven by independent Brownian motions and whose drift and volatility parameters depend on an independent Markov Chain Z, represent a simple example of a process $X = (Y^1, Y^2, Z)$ that satisfies Assumption 2.7.1. More generally, Duffie et al. (2003) show for the exponential affine case that if $Y^2 = 0$, Y^1 will be an Ornstein-Uhlenbeck type process. Although knowing that Y^1 is in general not an Ornstein-Uhlenbeck type process, we will refer to it in the following as the Ornstein-Uhlenbeck component of the process while we call Y^2 the non-negative component of X and Z its regime.

It is worthwhile to further investigate the implications of our specification for the

regime process Z. Assumption 2.7.1 entails that Z is a pure jump process with transition intensities. This means that $\vartheta(dt, dx_{d'})$ can be written as

$$\begin{aligned}\vartheta(dt, dx_{d'}) &= \sum_{n \neq Z(t-)} \varepsilon_{e_n - Z(t-)}(dx_d)\mu_n(t, X(t))dt, \qquad \text{where} \\ \mu_n(t, X(t)) &= Z(t)\chi_n^1(t),\end{aligned}$$

and $\mathcal{O}_n(t) - \int_0^t \mu_n(X(s))ds$ is an $\mathbf{F}$-martingale with $\mathcal{O}_n(t) := \sum_{0 \leq s \leq t} 1_{Z(s-) \neq e_n, Z(s) = e_n}$ denoting the point process that counts transitions of Z into state n. $\varepsilon_{e_n - e_m}(dx_d)$ denotes the positive point mass measure at $e_n - e_m$ of size one. Moreover, χ^1 denotes some matrix-valued function.

Although our setup constitutes an extension of the exponential affine setups usually considered in the literature, the next theorem shows that we obtain a comparable degree of complexity when calculating important transforms of the process $Y = (Y^1, Y^2)$:

Theorem 2.7.1 *Let $X = (Y^1, Y^2, Z)$ satisfy Assumption 2.7.1 and let $u(\cdot, T) : [0, T] \to \mathbb{C}^r$ and $v(\cdot, T) : [0, T] \to \mathbb{C}^{1 \times (d'-1)}$ be complex-valued deterministic functions solving the following system of ordinary differential equations (ODEs):[22] u solves*

$$\begin{aligned}0 = \dot{u} + v\beta^1_{(\Xi)} + \frac{1}{2}v\gamma^1_{(\Xi,\Xi)}v^T \\ + \int \left(e^{x_{d'}u + vx_{(\Xi)}} - 1\right) \nu^1(dx),\end{aligned} \tag{2.13}$$

and for each $l \in \{1, \ldots, d' - 1\}$ v_l solves

$$\begin{aligned}0 &= \dot{v}_l + v\beta^2_{(\Xi)l} \qquad l \leq d'_1 \\ 0 &= \dot{v}_l + v\beta^2_{(\Xi)l} + \frac{1}{2}v\gamma^2_{(\Xi,\Xi)(l-d'_1)}v^T + \int \left(e^{vx_{(\Xi)}} - 1\right) \nu^2_{(l-d'_1)}(dx_{(\Xi)}) \qquad l > d'_1\end{aligned} \tag{2.14}$$

with boundary conditions $v_l(T, T) = -c_l$ and $u_m(T, T) = 0$ for all $1 \leq m \leq r$. Furthermore, Ξ denotes the subset $\Xi := \{1, \ldots, d' - 1\}$. Then, the transform ${}_{T,t}\varphi_Y$ is given as

$${}_{T,t}\varphi_Y(c) := \mathbb{E}\left[e^{-c^T(Y(T) - Y(t))} \middle| \mathcal{F}'_t\right] = e^{Z(t)u(t,T) + \left(v(t,T) + c^T\right)Y(t)} \tag{2.15}$$

for

$$c = \begin{cases} -iw & , \quad w \in \mathbb{R}^{d'-1} \quad or \\ w & , \quad w \in \mathbb{R}^{d'-1}_{0+}, \end{cases}$$

[22] We suppress the dependence of β^1, β^2, γ^1, γ^2, ν^1 and ν^2 on t. Also, the dependence of u and v on t and the final time T is suppressed. The ODEs have to be understood as ODEs in t.

given that the following conditions are fulfilled:

$$\begin{aligned}
\infty &> \mathbb{E}\left[|H(T)|\right], \\
\infty &> \mathbb{E}\left[\int_0^T H(t)v\left(Z(t)\gamma^1(t)+\gamma^2(t)Y^2(t)\right)v^T H(t)dt\right], \qquad (2.16)\\
\infty &> \mathbb{E}\Bigg[\int_0^T \Big|H(t)\Big(\int\left(e^{vx_{(\Xi)}}-1\right)\nu^2(dx_{(\Xi)})Y^2(t) \\
&\qquad +Z(t)\int\left(e^{x_{d'}u+vx_{(\Xi)}}-1\right)\nu^1(dx)\Big)\Big|dt\Bigg], \qquad (2.17)
\end{aligned}$$

where H is defined as

$$H(t) := e^{Z(t)u(t,T)+v(t,T)Y(t)}.$$

Proof: For H defined as above, it follows from the Itô-formula that

$$\begin{aligned}
dH(t) &= H(t)\Bigg(\left(Z(t)\dot{u}+\dot{v}Y(t)\right)+v\left(Z(t)\beta^1_{(\Xi)}+\beta^2_{(\Xi)}Y(t)\right) \\
&\quad +\frac{1}{2}v\left(Z(t)\gamma^1_{(\Xi,\Xi)}+\left(\gamma^2Y^2(t)\right)_{(\Xi,\Xi)}\right)v^T \\
&\quad +\Bigg(\sum_{l=d_1'+1}^{d'-1}\left(\int\left(e^{vx_{(\Xi)}}-1\right)\nu^2_{(l-d_1')}(dx_{(\Xi)})Y_l^2(t)\right) \\
&\qquad +Z(t)\int\left(e^{x_{d'}u(t,T)+v(t,T)x_{(\Xi)}}-1\right)\nu^1(dx)\Bigg)\Bigg)dt \\
&\quad +H(t)v\left(\left(Z(t)\gamma^1_{(\Xi)}+(\gamma^2Y^2(t))_{(\Xi,\Xi)}\right)^{\frac{1}{2}}dW(t)\right) \\
&\quad +H(t-)\int\left(e^{x_{d'}u+vx_{(\Xi)}}-1\right)\tilde{J}(dt,dx),
\end{aligned}$$

where $\tilde{J}$ denotes the compensated jump process associated with the jumps of X. Since u and v solve the ODE system (2.13) and (2.14), the drift of H is 0. Furthermore, the second term is a martingale due to condition (2.16) and the last term is a martingale due to condition (2.17) (see e.g. Theorem 8 of Brémaud (1981), p. 27).
Hence, H is a martingale, i.e. $H(t) = E[H(T)|\mathcal{F}_t]$, and we have that

$$\mathbb{E}\left[e^{-c^TY(T)}\Big|\,\mathcal{F}_t\right] = e^{Z(t)u(t,T)+v(t,T)Y(t)},$$

which finally yields the claim.

□

The bottom line of Theorem 2.7.1 is the following: Given a process Y whose characteristics depend on the state of a finite state Markov process Z but are affine in Y

for each state of the regime process Z, the characteristic function (or equivalently the Laplace transform) of its increments can be calculated by solving a system of ODEs. The complexity of the problem therefore remains comparable to the one of the exponential affine setup, which is a special case of the setup considered here. However, in our case the ODE system in fact becomes a system of an ODE system – for each state of Z we obtain an extra ODE; the dimension of the overall system thus grows in the dimension r of the state space of the regime process Z.

If no particularly "nice" dependence of the characteristics of Y on (Y, Z) is presumed, we would generally obtain a Partial-Integro-Differential equation (PIDE) instead of the ODE system. The main advantage of obtaining an ODE system compared to a PIDE is that ODEs can numerically much easier be treated. Especially for model implementations, efficient ODE solvers are readily available in standard software.

It is important to note that our general result may be useful for a much broader class of applications than considered in this thesis. For example, it could also be employed to price plain-vanilla options on stock prices whose characteristics fit into the general setup of Theorem 2.7.1 since in this case the characteristic function of the stock price is given analytically up to the solution of equations (2.13) and (2.14). Based on the characteristic function, prices could then be obtained by Fourier inversion (see Carr and Madan (1999)). However, most important in the context of this thesis Theorem 2.7.1 will be our central tool for the computation of survival probabilities and the distribution of the loss process L in our applications of Chapters 3 and 4.

2.7.2 Calculation of Survival Probabilities and the Loss Distribution

Based on Theorem 2.7.1, we next consider the computation of survival probabilities and of the distribution of the loss process. Both represent quantities which are central to the applications in this thesis. In order to apply the theorem, we have to customize our stopping times model of Definition 2.1.1 such that Assumption 2.7.1 is fulfilled.

In Section 2.3, we have shown that in our model each N_i can be expressed in terms of orthogonal point processes $\left(N_1^*, N_2^*, \ldots, N_{\Pi'}^*, \ldots, N_{\{1,\ldots,I\}}^*\right)$. For all $\Pi' \subseteq \{1, \ldots, I\}$, the processes $N_{\Pi'}^*$ jump at most once and have compensators of the form

$$\int_0^t \lambda_{\Pi'}^*(s, X^1(s), X^2(s)) 1_{\tau_{(\Pi')} > s} ds.$$

The single jump processes N_i are reobtained as

$$N_i = \sum_{\Pi' \subseteq \{1,\dots,I\}} 1_{i \in \Pi'} N^*_{\Pi'}.$$

In order to calculate the survival probability of a subset Π of objects, we have to compute the following expectation (cf. Proposition 2.3.2):

$$\begin{aligned} {}_{T,t}p_\Pi &= {}^{\Pi}_{T,t}\varphi_{\Lambda_{(\Pi)}}(1) = \mathbb{E}^\Pi\left[\prod_{i\in\Pi} e^{-\Lambda_i(T)+\Lambda_i(t)}\middle|\mathcal{F}_t\right] = \\ &= \mathbb{E}^\Pi\left[e^{-\int_t^T \sum_{\Pi\cap\Pi'\neq\emptyset} \lambda^*_{\Pi'}(s)ds}\middle|\mathcal{F}_t\right] =: {}^{\Pi}_{T,t}\varphi_{\Gamma_{(\Psi)}}(1). \end{aligned} \tag{2.18}$$

Here, Ψ denotes a subset of the power set $\mathcal{P}(\{1,\dots,I\})$ and is defined as

$$\Psi := \{\Pi' \subseteq \{1,\dots,I\} : \Pi' \cap \Pi \neq \emptyset\},$$

and each coordinate $\Gamma_{\Pi'}$ of the $|\mathcal{P}(\{1,\dots,I\})|$-dimensional process Γ is given by the integrated intensity $\lambda^*_{\Pi'}$, i.e. we have

$$\Gamma_{\Pi'}(t) := \int_0^t \lambda^*_{\Pi'}(s)ds \qquad \forall t \in [0,T^*].$$

It is now important to note that the calculation of the expectation in equation (2.18) looks very much like a problem to which Theorem 2.7.1 can be applied, because we evaluate the joint Laplace transform of all integrated intensities which are related to a jump of $N_{(\Pi)}$. In fact, an application would be possible for any Π if the process

$$(X^1, X^2, N^*, \lambda', \Gamma),$$

describing the state of the whole system, satisfies Assumption 2.7.1. Here, λ' denotes the $|\mathcal{P}(\{1,\dots,I\})|$-dimensional process whose coordinates are given by the intensities $\lambda^*_{\Pi'}(t)1_{\tau_{(\Pi')}>t}$. Using the theorem, we could then semi-analytically calculate survival probabilities and based on these survival probabilities derive the loss distribution, too, as shown below.

In order to obtain a setup in which $(X^1, X^2, N^*, \lambda', \Gamma)$ is of the desired form, we have to consider a slight modification of our stopping times model. Namely, we have to presume that the compensator of each N^*_Π is given by

$$\int_0^t \lambda^*_\Pi(s, X^1(s), X^2(s))ds, \tag{2.19}$$

leaving everything else unchanged. Note that a similar modification has already been considered at the end of Subsection 2.6.2. The modification implies that the

N^*_Π are now "normal" point processes and not only one-jump point processes. The first jump is still associated with $1_{\forall i \neq j \in \Pi, l \notin \Pi:\ \tau_i = \tau_j \leq t, \tau_i \neq \tau_l}$ and the stopping times τ_i, but after the first jump the process remains "active" and may further jump and influence the other processes. Technically, we again have to impose condition (2.10) such that non-existing integrals are avoided.

In addition, we make the following assumptions:

1. The background process X^1 satisfies Assumption 2.7.1, i.e. can be written as $X^1 = (Y^1, Y^2, Z)$ with Y^1 the OU-type component, Y^2 the non-negative component and Z the regime process.

2. The contagion process X^2 evolves according to the stochastic differential equation (SDE)

$$dX^2(t) = \left(\chi^1 + \chi^2 X^2(t)\right) dt + \chi^3 dN^*(t)$$

with χ^1 being a constant vector and χ^2 and χ^3 being constant matrices such that X^2 is a non-negative process.

3. The intensities $\lambda^*_{\Pi'}$ are given as

$$\lambda^*_{\Pi'}(t) = (Z(t)\chi^{*1}(t) + \chi^{*2}(t)Y^2(t) + \chi^{*3}(t)X^2(t)),$$

where χ^{1*}, χ^{2*} and χ^{3*} denote positive vector-valued, vector-valued and matrix-valued functions of time.

One can easily check that under these assumptions, with

$$\begin{aligned} Y^{1\prime} &:= Y^1, \\ Y^{2\prime} &:= (Y^2, X^2, N^*, \lambda^*, \Gamma) \quad \text{and} \\ Z' &:= Z \end{aligned}$$

the process

$$(Y^{1\prime}, Y^{2\prime}, Z') = (X^1, X^2, N^*, \lambda^*, \Gamma) \tag{2.20}$$

satisfies Assumption 2.7.1.

Since we have that

$$\lambda_i(t) = \sum_{\Pi' \subseteq \{1,\ldots,I\}} 1_{i \in \Pi'} \lambda^*_\Pi(t),$$

our assumptions entail that the intensities of the N_i exhibit an affine dependence on Y^2 and X^2. In terms of Definition 2.1.1, the assumptions therefore correspond to a drift and a jump part of the jump-trigger process Λ, showing an affine dependence on Y^2 and X^2 for each state of Z, and a contagion process X^2, which is affine in

N^*.

It is, however, important to note that without the modification made in (2.19) the process describing the state of the system is not of the desired form because without modification the intensities of N^*_Π are given by $1_{\tau_{(\Pi')}>t}\lambda^*_\Pi(t)$. They are therefore not affine in N^*_Π. However, the modification will only be necessary as long as the N^*_Π can influence the λ^*_Π. In a conditional independence setup, i.e. if the σ-fields $\bigvee_{i\in\{1,\dots,I\}}\sigma(E_i)$ and $\mathcal{F}^\Lambda_{T^*}$ are independent (see Proposition 2.5.1 and Definition 2.5.2), the modification can be omitted. In this case, conditional survival probabilities only depend on the process

$$(Y^{1\prime}, Y^{2\prime}, Z') = (Y^1, (Y^2, \lambda^*, \Gamma), Z), \tag{2.21}$$

which already satisfies Assumption 2.7.1 if the first and the third of our three basic assumptions with $\chi^{*3}(t) \equiv 0$ are fulfilled.

Calculation of Survival Probabilities

Depending on whether we work in a conditional independence setup or in a model with contagion, we now assume that $(Y^{1\prime}, Y^{2\prime}, Z')$ is either of the form given by equation (2.21) or by equation (2.20) and satisfies Assumption 2.7.1. Then, we obtain for the survival probability

$$_{T,t}p_\Pi = \mathbb{E}^\Pi\left[e^{-\sum_{\Pi'\cap\Pi\neq\emptyset}\int_t^T\lambda^*_{\Pi'}(s)ds}\middle|\mathcal{F}_t\right] = e^{Z'(t)u(t,T)+v(t,T)Y'(t)},$$

where u and v solve equations (2.13) and (2.14) with appropriate parameters c, β^1, β^2, γ^1, γ^2, ν^1 and ν^2.[23] While in the contagion case the transform is calculated under $\mathbb{P}^\Pi$, for a conditional independence model it is computed under the original measure $\mathbb{P}$. In this case, the dynamics of λ^* are the same under $\mathbb{P}^\Pi$ and $\mathbb{P}$.

All in all, Theorem 2.7.1 therefore presents a very useful tool based on which survival probabilities for a large range of model specifications can conveniently be computed. It is worth mentioning that finite state Markov processes have already been proposed for modeling the term structure of interest rates, see e.g. Landén (2000) and Bansal and Zhou (2002).[24] In particular, the Landén (2000) setup can be considered as a one-dimensional, special case of the setup considered here.

[23] In Collin-Dufresne et al. (2004), a similar result for the special case of the single survival probability and a purely affine setup is derived.

[24] Observe that the computation of the expectation in (2.18) comes close to the calculation of zero-coupon prices in an interest rate setup.

Calculation of the Loss Distribution

Apart from the survival probabilities, the most important quantity to be computed in applications is the loss distribution. Assuming that $\Pi \subseteq \{1, \dots, I\}$ objects have survived until t and all N_j for $j \notin \Pi$ have already jumped, we obtain for the loss distribution with $0 \leq n \leq |\Pi|$

$$\begin{aligned}\mathbb{P}\left(L(T) = n + I - |\Pi| \,\middle|\, \mathcal{F}_t \wedge \bigcap_{i\in\Pi}\{\tau_i > t\} \wedge \bigcap_{i\notin\Pi}\{\tau_i \leq t\}\right) \\ = \sum_{\Pi'\subseteq\Pi:|\Pi'|=|\Pi|-n} \sum_{k=0}^{|\Pi|-|\Pi'|} (-1)^k \sum_{\Pi''\subseteq\,\Pi\setminus\Pi':|\Pi''|=k} {}_{T,t}p_{\Pi'\cup\Pi''},\end{aligned}$$

where we set ${}_{T,t}p_{\emptyset} = 1$.[25] Here, the first sum runs over all events where exactly $|\Pi| - n$ objects survive, and the second and third sum compute the corresponding probability of this event. In the case of a *homogeneous* portfolio (i.e. if ${}_{T,t}p_{\Pi'} = {}_{T,t}p_{\Pi''}$ for all $\{\Pi'', \Pi' \subseteq \Pi : |\Pi'| = |\Pi''|\}$), the formula collapses to

$$\begin{aligned}\mathbb{P}\left(L(T) = n + I - |\Pi| \,\middle|\, \mathcal{F}_t \wedge \bigcap_{i\in\Pi}\{\tau_i > t\} \wedge \bigcap_{i\notin\Pi}\{\tau_i \leq t\}\right) \\ = \binom{|\Pi|}{|\Pi| - n} \sum_{k=0}^{n} \binom{n}{k} (-1)^k \, {}_{T,t}p_{\{1,\dots,k+|\Pi|-n\}}.\end{aligned}$$

Here, ${}_{T,t}p_{\{1,\dots,k+|\Pi|-n\}}$ is defined as the survival probability of the first $k + |\Pi| - n$ objects, which is the same for any subset containing $k + |\Pi| - n$ objects due to the presumed homogeneity.

If the process given by equation (2.20) satisfies Assumption 2.7.1, we can compute the survival probabilities ${}_{T,t}p_{\{1\}}, \dots, {}_{T,t}p_{\{1,\dots,n\}}$ and derive the portfolio loss distribution. However, this procedure works only in principle for arbitrary portfolio sizes I. While in the heterogeneous case the number of summands quickly exceeds computational resources as I grows, in the homogeneous case we face the problem that for large I we have to multiply large values like $\binom{|\Pi|-n}{k}$ with typically small values like ${}_{T,t}p_{\{1,\dots,k\}}$ generating large error terms.[26]

The described computation of the loss distribution based on survival probabilities is possible for a contagion as well as a conditional independence setup, but it is limited to portfolios consisting only of few objects. Contrarily, the technique which we explain next is feasible for arbitrary portfolio sizes but strongly depends on the

[25] We assume that $\emptyset \in \Pi$ for any subset Π of $\{1, \dots, I\}$.

[26] In numerical experiments, we found that the described approach typically loses its computational tractability for portfolio sizes of about 15 objects depending on the respective model parametrization.

conditional independence assumption; it goes back to an algorithm introduced by Andersen et al. (2003) and has been applied the first time to a dynamic setup by Mortensen (2006). As previously pointed out in this subsection, in order to apply Theorem 2.7.1 in the conditional independence case it is sufficient that the process given by equation (2.21) satisfies Assumption 2.7.1.

In Section 2.5, we have shown that the conditional independence setup eventually corresponds to a model-implied dependence structure which is characterized by a copula of the Marshall-Olkin type. As pointed out there, it is convenient to assume a certain factor structure for the coordinates of the jump-trigger process Λ such that the dependence structure simplifies to a LT-Archimedean copula. Following Section 2.5, we denote this p-dimensional factor process by Λ^*.

As a result, conditional on the evolution of this factor the jumps of N are independent, which is at the bottom of the Andersen et al. (2003) / Mortensen (2006) approach. It exploits that the survival probability of each object i conditional on the factor realization x can be written as

$$ {}_{T,t}p_i\,(x) := \mathbb{P}\left(\tau_i > T \middle|\, \mathcal{F}_t \wedge \{\Lambda^*(T) - \Lambda^*(t) = x\} \wedge \{\tau_i > t\}\right). $$

Using the conditional independence, we then obtain[27]

$$ \begin{aligned} &\mathbb{P}\left(\sum_{i=1}^{K} 1_{\tau_i \le T} = n \middle| \mathcal{F}_t \wedge \{\Lambda^*(T) - \Lambda^*(t) = x\} \wedge \bigcap_{i=1}^{I}\{\tau_i > t\}\right) = \\ &\quad \mathbb{P}\left(\sum_{i=1}^{K-1} 1_{\tau_i \le T} = n \middle| \mathcal{F}_t \wedge \{\Lambda^*(T) - \Lambda^*(t) = x\} \wedge \bigcap_{i=1}^{I}\{\tau_i > t\}\right) {}_{T,t}p_K\,(x) \\ &\quad + (1 - {}_{T,t}p_K\,(x))\,\mathbb{P}\left(\sum_{i=1}^{K-1} 1_{\tau_i \le T} = n-1 \middle| \mathcal{F}_t \wedge \{\Lambda^*(T) - \Lambda^*(t) = x\} \wedge \bigcap_{i=1}^{I}\{\tau_i > t\}\right). \end{aligned} \tag{2.22} $$

For $K = I$, this finally yields the conditional loss distribution and the loss distribution itself by integrating the p-dimensional factor:

$$ \begin{aligned} &\mathbb{P}\left(L(T) = n \middle| \mathcal{F}_t \wedge \{\Lambda^*(T) - \Lambda^*(t) = x\} \wedge \bigcap_{i=1}^{I}\{\tau_i > t\}\right) \\ &\qquad = \int \mathbb{P}\left(L(T) = n \middle| \mathcal{F}_t \wedge \{\Lambda^*(T) - \Lambda^*(t) = x\} \wedge \bigcap_{i=1}^{I}\{\tau_i > t\}\right) d\,{}_{T,t}G_{\Lambda^*}\,(x), \end{aligned} \tag{2.23} $$

where ${}_{T,t}G_{\Lambda^*}$ denotes the distribution function of $\Lambda^*(T) - \Lambda^*(t)$ conditional on $\mathcal{F}_t$.

Given that the process $(Y^{1\prime}, Y^{2\prime}, Z')$ defined by equation (2.21) satisfies Assumption

[27] To keep notation simple, we now assume that all I objects have survived until t. The case where only a subset Π has survived can be treated analogously.

2.7.1, the characteristic function of the independent coordinates $\Lambda_j^*(T) - \Lambda_j^*(t)$ can semi-analytically be calculated by applying Theorem 2.7.1, and afterwards inverted to get the corresponding density via

$$ {}_{T,t}g_{\Lambda_j^*}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwx} \, {}_{T,t}\varphi_{\Lambda_j^*}(-iw) dw. \tag{2.24} $$

In order to keep the computation of the multi-dimensional integral in equation (2.23) tractable, one usually assumes a low-dimensional factor structure. For instance, in Mortensen (2006) a one-factor structure of the Duffie and Gârleanu (2001) credit portfolio model is considered.

It is worth mentioning that in the conditional independence setup, also a limiting loss distribution can be derived (for details see Jarrow et al. (2005)). Namely, it is the distribution that is obtained by letting $I \to \infty$. For example, under the homogeneity assumption one obtains

$$ \frac{1}{I} L \quad \longrightarrow \quad 1 - {}_{T,t}p\left(\Lambda^*(T) - \Lambda^*(t)\right) \qquad\qquad (I \to \infty). \tag{2.25} $$

Convergence is in the L^2 sense (see Jarrow et al. (2005)). Using the limiting loss distribution, the recursive calculation of the conditional loss distribution from equation (2.22) can be avoided. In addition, for very simple model specifications (distributions of $\Lambda^*(T) - \Lambda^*(t)$), analytical formulas for the prices of plain vanilla options on the loss can be derived.

We end this section by briefly introducing a third possibility of calculating the loss distribution which arises from the fact that, under the assumption that X^1 and X^2 are Markov processes, the loss process itself is a finite state Markov process of state space dimension $2^{|I|}$. For an investigation of this link see e.g. Schönbucher (2005), Herbertsson and Rootzén (2006) or Frey and Backhaus (2007). Under the homogeneity assumption, the state space dimension reduces to $(I+1)$ since merely the number of already jumped processes is relevant and the single processes do not have to be identified in order to determine the probabilities of further jumps.

Under the (rather restrictive) assumption that the generator matrix $A(t, X^1(t))$ associated with the finite state Markov process L can be decomposed as[28]

$$ A(t, X^1(t)) = V \, \mathrm{diag}(-\alpha_1(X^1(t)), \ldots, -\alpha_{D-1}(X^1(t)), 0) \, V^{-1} $$

with α_d being positive functions of X^1, the transition probabilities of L are given as (see e.g. Lemma 1 of Hurd and Kuznetsov (2007), p. 10):

$$ \mathbb{P}(L(T) = j | L(t) = i) = \sum_{d=1}^{D} v_{id} \tilde{v}_{dj} \mathbb{E}\left[e^{-\int_t^T \alpha_i(X^1(s)) ds} \,\middle|\, \mathcal{F}_t \right]. \tag{2.26} $$

[28]Since the last state is usually associated with the situation where all objects have jumped and therefore is an absorbing state, the last row of A only contains zeros.

Here, v_{id} and $\tilde{v}_{dj}$ denote the entries of the matrices V and V^{-1}. Let now $X^1 = (Y^1, Y^2, Z)$ satisfy Assumption 2.7.1 and the α_d be affine in Y^2 for each state of the regime Z. Then, the expectation in equation (2.26) – and thus the distribution of L – can semi-analytically be computed by applying Theorem 2.7.1. This approach has been introduced by Lando (1998) to model a firm's rating transitions and has been also used by Hurd and Kuznetsov (2007).

2.8 Towards Applications: Credit Portfolio Risk and Stochastic Mortality Modeling

After having analyzed the dependence structure of our general stopping times model introduced in Definition 2.1.1 and having addressed the issue of model tractability, this section provides a link between the general setup and the more specific models considered for the applications of Chapters 3 and 4. Namely, we will apply our stopping times setup to dynamically model credit portfolios in Chapter 3 and to price and analyze catastrophe mortality linked securities in Chapter 4.

2.8.1 Credit Portfolio Risk Modeling

Credit portfolio modeling is an important task for banks and other financial institutions which have credit-sensitive securities on their balance sheets. Broadly speaking, the credit risk associated with each single object is the risk that an object's ability to pay back its financial obligations might substantially deteriorate. As a result, its probability to default could increase or the company might even default. To assess and control these risks, credit portfolio models are needed which are able to account for the different risk drivers of a credit portfolio.

In addition, since the mid 90s banks have begun to actively manage their credit risks by means of securitization, i.e. by repackaging single risks and selling them to investors. The market for such structured credit products has shown tremendous growth during recent years (see e.g. Duffie (2007)). By construction, the securities reference a whole portfolio and not merely one single object anymore. In order to price these securities or assess their risk, the dynamics of the whole underlying portfolio have to be modeled.

As already pointed out at the beginning of this chapter, our stopping times model of Definition 2.1.1 belongs to the reduced-form model class because we model the default times as inaccessible stopping times (see Section 2.2). It represents an ideal candidate for modeling portfolio credit risks since it can incorporate many different phenomena. Namely, default probabilities of firms can depend on past defaults of others through the dependence of the default-trigger processes Λ_i on N. On the

other hand, the background process X^1, on which the Λ_i also depend, can model firm-specific risk drivers or macroeconomic sources of risk such as interest rates, which may have an influence on defaults. Also, simultaneous defaults are possible in the model because the Λ_i are allowed to jump together.

Reduced-form credit portfolio models can be further subdivided into *Bottom-Up* and *Top-Down* models. Bottom-Up means that one specifies a model for the joint dynamics of the single processes N_i. Based on this specification, one then derives the dynamics of the aggregate process $L = \sum_{i=1}^{I} N_i$. Contrarily, in the Top-Down approach one directly specifies a model of L. While our setup comes closer to the Bottom-Up approach, we can also choose the Top-Down perspective as shown in Section 2.6.

In order to illustrate the usefulness of the theoretical tools developed in this chapter, in the context of a credit portfolio application, we apply these next to analyze the Duffie and Gârleanu (2001) model; it can be considered as the standard model within the reduced-form modeling approach due to its popularity and many appealing features (see Mortensen (2006), Feldhütter (2008) and Eckner (2007) for empirical investigations of the model).

Example: Analysis of the Duffie and Gârleanu (2001) Model

Basically, the Duffie and Gârleanu (2001) model aims at modeling single and portfolio credit risks at the same time, i.e. represents a Bottom-Up model, and relies on the conditional independence assumption (see Subsection 2.5.1). In terms of our stopping times model, this means that the default-trigger process Λ evolves independently of past defaults. Duffie and Gârleanu (2001) further propose a factor structure for the Λ_i such that

$$\Lambda_i(t) = \sum_{j=1}^{p} \int_0^t \lambda_j^c(s)ds + \int_0^t \tilde{\lambda}_i(s)ds.$$

In terms of Definition 2.1.1:

- The background process X^1 is given by

$$X^1 = \left(\lambda_1^c, \ldots, \lambda_p^c, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_I\right).$$

  Its coordinates are assumed to be positive processes, which evolve mutually independently (for details see below), and $\mathcal{F}_{T^*}^{X^1} = \mathcal{F}_{T^*}^{\Lambda}$ and $\mathcal{F}_t^N$ are independent given $\mathcal{F}_t^{X^1}$.

- $b_i(s, X^1(s), X^2(s)) = \sum_{j=1}^{p} \lambda_j^c(s) + \tilde{\lambda}_i(s)$.

- $\nu_m(s, \mathbb{R}^I_{0+}) = 0$, for each m.

Since $\nu_m(s, \mathbb{R}^I_{0+}) = 0$, simultaneous defaults are not possible in the Duffie and Gârleanu (2001) model (cf. Proposition 2.3.1).

The interpretation of the λ^c_j is the one of common risk factors which all firms are exposed to; the $\tilde{\lambda}_i$ represent firm-specific risks that are in the credit portfolio risk terminology called idiosyncratic risks. In Duffie and Gârleanu (2001), each λ^c_j and $\tilde{\lambda}_i$ evolves according to the SDE

$$d\lambda(t) = \kappa\left(\eta - \lambda(t)\right)\, dt + \sigma\sqrt{\lambda(t)}\, dW(t) + dJ(t), \quad \lambda(0) = \overline{\lambda}. \tag{2.27}$$

Here κ, σ and η are positive constants, W is a standard Wiener process and J a Poisson process that jumps with intensity μ and has positive, independently $\mathrm{Exp}\left(\frac{1}{\zeta}\right)$-distributed jumps. Following Duffie and Gârleanu (2001), we call a process of the form (2.27) a *basic affine jump diffusion* and denote it by

$$BAJD\left(\overline{\lambda}_0, \eta, \kappa, \sigma, \mu, \zeta\right).$$

For ease of exposition, we restrict ourselves in the following model analysis to the parsimonious model specification of Mortensen (2006), which constitutes a slight modification of the original model. This modification is usually considered in empirical investigations of the model (see Mortensen (2006), Feldhütter (2008) and Eckner (2007)). We would like to mention that the following results carry over to the general case, but the notation becomes substantially more involved without leading to further insights. In Mortensen (2006), a one-factor version of the model is considered where a positive constant determines how strong each firm is exposed to the common factor. More precisely,

$$\Lambda_i(t) = \int_0^t a_i \lambda^c(s) ds + \int_0^t a_i \tilde{\lambda}_i(s) ds,$$

where $a_i > 0$ for all $i \in \{1, \ldots, I\}$ and

$$\begin{aligned} \lambda^c \quad & \text{is a} \quad BAJD\left(\overline{\lambda}^c, \varpi\eta, \kappa, \sigma, \varpi\mu, \zeta\right), \\ \tilde{\lambda}_i \quad & \text{is a} \quad BAJD\left(\overline{\lambda}_i, (1-\varpi)\eta, \kappa, \sigma, (1-\varpi)\mu, \zeta\right) \end{aligned}$$

and $\varpi \in [0, 1]$ is a constant. Note that ϖ controls the level of dependence between the portfolio objects. Assume, for example, $\varpi = 1$. Then, jumps of the default intensities are purely systematic, i.e. the jump intensities always jump at the same time and $\tilde{\lambda}_i$ never jumps in this case. However, jump sizes can still be different due to different a_i values. Contrarily, if $\varpi = 0$ intensities will never jump together because λ^c never jumps. Therefore, ϖ can be interpreted as the weight of the

common factor in the intensity specification.[29]

An application of the Itô-formula, with λ^c and $\tilde{\lambda}_i$ being independent basic affine jump diffusions with the same jump intensity and speed of mean reversion yields that

$$a_i\left(\lambda^c(s)+\tilde{\lambda}_i(s)\right) \qquad \text{is a} \qquad BAJD\left(a_i\left(\overline{\lambda}^c+\overline{\lambda}_i\right)a_i\eta,\kappa,\sqrt{a_i}\sigma,\mu,a_i\zeta\right),$$

i.e. that the jump intensity μ and the speed of mean reversion κ of all default intensities are the same. Furthermore, their mean reversion level $a_i\eta$, their volatility $\sqrt{a_i}\sigma$ and their jump size mean $a_i\zeta$ depend on common values but are scaled by a positive constant a_i.[30]

As indicated in Section 2.5, in the model calibration conducted by Mortensen (2006) the a_is are used to match individual risk levels, while the remaining parameters are primarily adjusted to fit dependencies between the single risks. Since the model does not exhibit contagion, the model's dependence function is a copula of the Marshall-Olkin type (cf. Theorem 2.5.2); single survival probabilities can further be written as

$${}_{T,t}p_i = \mathbb{E}\left[e^{-a_i\left(\sum_{j=1}^{I+1}A_{ij}\left(\Lambda_j^*(T)-\Lambda_j^*(t)\right)\right)}\middle|\,\lambda(t)\right],$$

where

$$\begin{aligned}\Lambda_1^*(t) &:= \int_0^t \lambda^c(s)ds\\ \Lambda_j^*(t) &:= \int_0^t \tilde{\lambda}_{j-1}(s)ds, \quad 2\le j\le I+1,\end{aligned}$$

and $A\in\mathbb{R}^{I\times(I+1)}$ is of the form

$$A=\begin{pmatrix}1&1&0&\cdots&0\\1&0&1&\cdots&0\\\vdots&\vdots&&\ddots&\vdots\\1&0&\cdots&\cdots&1\end{pmatrix}.$$

From Corollary 2.5.1 it therefore follows that the model-implied dependence structure is characterized by a LT-Archimedean copula function (with factors $\Lambda^*(T)-\Lambda^*(t)$ and a matrix A as above), which we want to further investigate next.

To keep calculations simple, we set $\overline{\lambda}_i=0$ for all $i\in\{1,\ldots,I\}$ and $\varpi=1$ implying

[29] Note that ϖ does not only determine the degree of common jumps, it also determines the mean reversion levels of factor and idiosyncratic components and therefore influences the Brownian components of the processes, too.

[30] In Proposition 1 of Duffie and Gârleanu (2001), an alternative derivation of this result can be found based on the Laplace transforms of the processes.

that $\tilde{\lambda}_i(s) \equiv 0$ for all $s \geq 0$; under this assumption, the model-implied dependence structure is given by an ordinary Archimedean copula function (cf. Corollary 2.5.1) where φ of equation (2.6) is given by

$$\varphi(u) := {}_{T,t}\varphi_{\int_0^\bullet \lambda^c(s)ds}(u) = \mathbb{E}\left[e^{-u\int_t^T \lambda^c(s)ds} \middle| \lambda(t)^c \right].$$

Fortunately, this Laplace transform can be calculated analytically (cf. Appendix A) facilitating the calculation of dependence measures such as Kendall's tau.[31] Figure 2.1 displays Kendall's tau for different initial intensity states $\overline{\lambda}^c$.[32] We find that the dependence level decreases in the initial intensity state, which might be surprising at first sight since high $\overline{\lambda}^c$ values imply a higher risk for all portfolio objects. However, from a dependence perspective the randomness of the integrated process $\int_t^T \lambda^c(s)ds$ alone is the critical quantity. For higher initial intensity states $\overline{\lambda}^c$, the joint evolution of future intensities becomes less random since high intensities are almost deterministically pulled down towards the mean reversion level η. In this regard, note that if $\int_t^T \lambda^c(s)ds$ was deterministic one would obtain the independence copula as a limiting case.

Our findings are confirmed by Figure 2.2, where we present 2000 simulations from the copulas corresponding to initial intensity states of $\overline{\lambda}^c = 0.2$ and $\overline{\lambda}^c = 0.0002$, keeping the other parameters fixed. While for $\overline{\lambda}^c = 0.2$ realizations seem to be almost equally distributed, in case of $\overline{\lambda}^c = 0.0002$ dependence is visible as points are more clustered around the 45° line.

We now focus on the characterization of the model-implied default clustering from a dynamic point of view. Since simultaneous defaults occur with probability 0 in the considered model, the compensator of the loss process equals the compensator of the quadratic variation of the loss process (cf. Section 2.6); it is given by

$$\langle L, L\rangle(t) = \sum_{i=1}^{I} \Lambda_i(t) = \int_0^t \left(\sum_{i=1}^{I} a_i 1_{\tau_i > s} \right) \lambda^c(s)ds + \sum_{i=1}^{I} \int_0^t 1_{\tau_i > s} a_i \tilde{\lambda}_i(s)ds.$$

The expected volatility (cf. Definition 2.6.1) associated with the model is therefore

[31] Nevertheless, computation has to be done numerically. The lacking smoothness of the graph in Figure 2.1 has to be attributed to numerical errors.

[32] For illustration purposes we used the model parametrization estimated in Mortensen (2006) with the only differences that $\varpi = 1.0$ and $\overline{\lambda}_i = 0$. However, since $\varpi = 0.91$ in Mortensen (2006) differences should be minor.

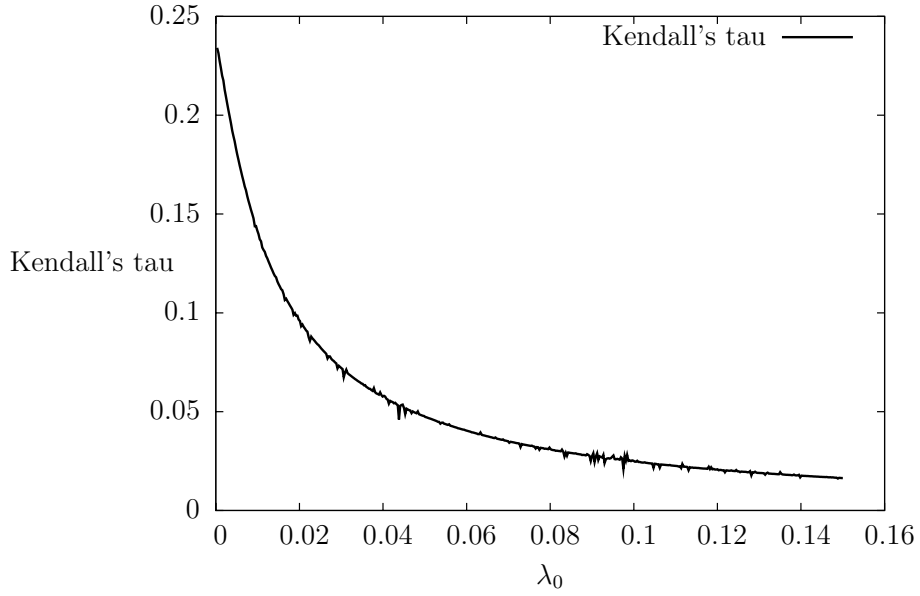


Figure 2.1: Kendall's tau with respect to different intensity states $\overline{\lambda}^c$ for the five-year horizon in the Duffie and Gârleanu (2001) model. Remaining parameters are fixed at $\kappa = 0.27$, $\varpi = 1$, $\eta = 0.0046$, $\sigma = 0.05$, $\mu = 0.017$, $\zeta = 0.078$ and $\overline{\lambda}_i = 0$.

$$
\begin{aligned}
{}_{T,t}\sigma_L &= \mathbb{E}\left[\left[L,L\right](T) - \left[L,L\right](t)\middle|\,\mathcal{F}_t\right] \\
&\quad + \mathbb{E}\left[\left(\left[\sum_{i=1}^{I} a_i 1_{\tau_i>\bullet}\left(\lambda^c(\bullet)+\tilde{\lambda}_i(\bullet)\right), \sum_{i=1}^{I} a_i 1_{\tau_i>\bullet}\left(\lambda^c(\bullet)+\tilde{\lambda}_i(\bullet)\right)\right](T)\right.\right. \\
&\quad \left.\left. - \left[\sum_{i=1}^{I} a_i 1_{\tau_i>\bullet}\left(\lambda^c(\bullet)+\tilde{\lambda}_i(\bullet)\right), \sum_{i=1}^{I} a_i 1_{\tau_i>\bullet}\left(\lambda^c(\bullet)+\tilde{\lambda}_i(\bullet)\right)\right](t)\right)\middle|\mathcal{F}_t\right].
\end{aligned}
$$

By simple transformations, we obtain

$$
\begin{aligned}
{}_{T,t}\sigma_L &= \mathbb{E}\left[\sum_{i=1}^{I}\int_t^T a_i 1_{\tau_i>s}\left(\lambda^c(s)+\tilde{\lambda}_i(s)\right)ds\middle|\,\lambda(t)\right] \\
&\quad + \mathbb{E}\left[\int_t^T\left(\sum_{i=1}^{I} 1_{\tau_i>s}a_i\right)^2\sigma^2\lambda^c(s)ds + 2\zeta^2\varpi\mu\int_t^T\left(\sum_{i=1}^{I} 1_{\tau_i>s}a_i\right)^2 ds\middle|\,\lambda^c(t)\right] \\
&\quad + \sum_{i=1}^{I}\mathbb{E}\left[\int_t^T 1_{\tau_i>s}a_i^2\sigma^2\tilde{\lambda}_i(s)ds + 2\zeta^2(1-\varpi)\mu\int_t^T 1_{\tau_i>s}a_i^2 ds\middle|\,\tilde{\lambda}_i(t)\right] \\
&\quad + \mathbb{E}\left[\int_t^T\sum_{i=1}^{I} 1_{\tau_i>s}a_i^3\left(\lambda^c(s)+\tilde{\lambda}_i(s)\right)^3 ds\middle|\,\lambda(t)\right], \qquad (2.28)
\end{aligned}
$$

which follows from a calculation of the quadratic variation of the Brownian component and the jump part of $\lambda := (\lambda^c, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_I)$ exploiting the independence of its coordinates and the independence of λ from L. Moreover, due to the Markovian model structure, $\mathcal{F}_t^{X^1}$ can be replaced by $\lambda(t)$.

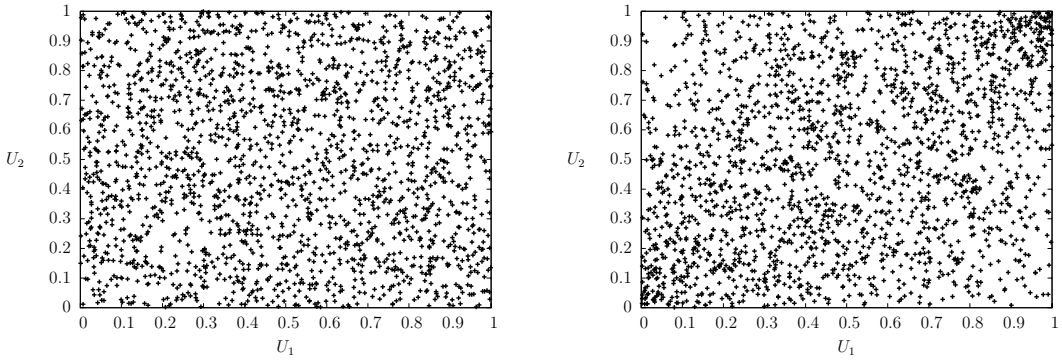


Figure 2.2: 2000 simulations from the copula implied by the Duffie and Gârleanu (2001) model with initial intensity state $\overline{\lambda}^c = 0.2$ (left) and $\overline{\lambda}^c = 0.0002$ (right). Remaining parameters are fixed at $\kappa = 0.27$, $\varpi = 1$, $\eta = 0.0046$, $\sigma = 0.05$, $\mu = 0.017$, $\zeta = 0.078$ and $\overline{\lambda}_i = 0$. The considered time horizon is five years.

The second term of equation (2.28) represents the contribution of the common factor λ^c to the expected volatility of the loss process, the third term contains the contributions of the idiosyncratic risks, while the last term includes contributions of defaults. We find that expected volatility in the Duffie and Gârleanu (2001) model increases *ceteris paribus* in ϖ, i.e. the measure grows with an increasing weight ϖ of the common factor since

$$\left(\sum_{i=1}^{I} a_i\right)^2 > \sum_{i=1}^{I} a_i^2.$$

The intensity of the loss process and the defaults become more volatile as the single risks are more and more exposed to systematic risk, represented by the factor λ^c. Furthermore, since for $a_{I+1} > 0$

$$\left(\sum_{i=1}^{I+1} a_i\right)^2 - \sum_{i=1}^{I+1} a_i^2 > \left(\sum_{i=1}^{I} a_i\right)^2 - \sum_{i=1}^{I} a_i^2,$$

the *relative* weights between the common factor and the idiosyncratic components in the calculation of the expected volatility are moved towards the common factor as the number of portfolio objects gets bigger. This is in concordance with a result of Vasicek (1991) who shows in a static setting that the idiosyncratic components

"disappear" in the calculation of the loss distribution as the number of portfolio objects increases to infinity (note that similar considerations have lead to equation (2.25)).

We now focus on the dependence structure by considering only the second and the third term of equation (2.28) and disregard effects on expected volatility that stem from the downward jumps of the aggregated intensity at defaults.[33] We find that in the Duffie and Gârleanu (2001) model both terms, considered as functions in $\lambda(t)$, evolve relatively stable over time since

$$\mathbb{E}\left[2\zeta^2\varpi\mu\int_t^T\left(\sum_{i=1}^I a_i\right)^2 ds + 2\zeta^2(1-\varpi)\mu\int_t^T\sum_{i=1}^I a_i^2 ds\,\middle|\,\lambda(t)\right] = const.$$

Only the contributions due to the Brownian components in the intensity specification depend on λ, but these contributions will usually be small since σ is typically small. Therefore, the Duffie and Gârleanu (2001) model comes close to a model where the default clustering does not vary over time.[34]

In summary, our examination of the Duffie and Gârleanu (2001) model yields that the model is able to generate realistic degrees of dependence, but for typical parametrizations the model comes close to a time-homogeneous model with respect to default clustering. Our results confirm the findings of an empirical study by Feldhütter (2008). He investigates the model's ability to capture the price dynamics of structured credit products. Following our theoretical reasoning, it is not surprising that he finds that the model cannot capture the price variation of the securities which are most exposed to changes in the dependence structure of the underlying portfolio – the so-called *senior tranches*. Without going into details at this point (for such details see Chapter 3), the riskiness of such senior tranches increases as the dependencies among the portfolio objects grow. Feldhütter (2008) further finds that the tranches' risk, represented by their prices, is negatively correlated with the state of the common factor λ^c. This again is no surprise since the model-implied dependence structure decreases in the state of λ^c as our analysis of the model-implied copula has demonstrated. Put differently, dependence structure and common factor cannot be positively correlated – at least for reasonable model parametrizations.

Our findings show that the developed tools represent valuable instruments for the analysis of credit portfolio models.

[33] This is not a too restrictive assumption since when analyzing the price dynamics of structured credit derivatives such as Itraxx CDO tranches usually no defaults are observed.

[34] Of course, this only holds true for parametrizations where σ is small and the jump part represents a significant part of the risk. However, these are the parametrizations which are typically obtained when calibrating the model to prices of structured credit products, see e.g. Mortensen (2006) and Feldhütter (2008).

2.8.2 Stochastic Mortality Modeling

The second application of our stopping times model introduced in this chapter are the valuation and risk analysis of catastrophe mortality linked securities considered in Chapter 4: We introduce a new model for analyzing and pricing these securites. The current subsection therefore provides a link between our general setting and the more specific setup considered for the application in an insurance related field. In particular, we introduce and formulate the quantities in terms of Definition 2.1.1, which will be of particular importance in Chapter 4.

One of the primary sources of risk that a (re)insurance company selling annuities or life insurance contracts faces is mortality risk. Broadly speaking, mortality risk is the possibility of future, systematic deviations of mortality rates from expected rates. In this context, one usually distinguishes between *longevity risk* and *catastrophe mortality risk*. While longevity risk is the risk – from an insurance company's point of view – that customers who bought annuities systematically live longer than expected, catastrophe mortality risk corresponds to the risk that mortality rates suddenly increase leading to a large payout in connection with the insurer's life business.

In order to assess mortality risk, one needs to specify a model of the future evolution of mortality rates as a first step. The first contributions on stochastic mortality modeling have followed approaches different from the one introduced in this subsection. By replacing formerly constant parameters in the Gompertz mortality law, Milevsky and Promislow (2001) were among the first to propose a *stochastic hazard rate* or spot force of mortality.[35]

Since then, there have been several articles utilizing ideas from credit risk modeling for the modeling of stochastic mortality rates.[36] We use an approach similar to the one of Miltersen and Persson (2005), which is based on the so-called intensity-based approach from Lando (1998) and arises as a special case of the setup considered in Proposition 2.5.1. For modeling the mortality rates of a portfolio of insureds, we now associate the death times of the insureds with the stopping times $\tau = (\tau_1, \ldots, \tau_I)$ introduced in Definition 2.1.1 – similar to the default times of firms in the credit portfolio setup discussed in the previous subsection.

However, there are some points that are special about an application in an insurance context: Consider a portfolio of insureds all having the same age and a

[35] *Mortality laws* are certain parametric functions which are used to model human mortality; see Gompertz (1825) for a classical example or Bowers et al. (1997) for an overview on mortality laws.

[36] Even earlier, Artzner and Delbaen (1995) pointed out similarities between "default risk insurance" and life insurance.

particular mortality rate. By assuming a large enough portfolio (more precisely an infinitely large portfolio), it suffices to model only the mortality rate and the actual realizations of the jumps, i.e. the deaths of the insureds, become irrelevant (cf. 2.25). Our reasoning further implies that when pricing a mortality contingent security depending on the evolution of mortality rates of such a large portfolio we should only attach a risk premium to the risk factors driving the mortality rate. There should be no compensation for taking the risk of the single jumps of N, i.e. the deaths, because this risk can be diversified away.

Nevertheless, there may be other situations where such a risk premium is justified. For example, it may be justified when considering only small cohorts. For a general discussion of this issue, see Bauer and Russ (2006). Since we want to apply the framework in Chapter 4 for modeling securities which depend on the mortality rates of very large cohorts (such as the male population in the U.S.), the following assumptions seem to be justified:

Assumption 2.8.1

1. *People who die do not influence the mortality rates of the surviving people.*
2. *People with the same age have the same survival probabilities over the same future time period.*

Translated into the modeling context of Definition 2.1.1, our assumptions imply that a stochastic mortality model should not exhibit contagion (cf. Definition 2.5.2) and that we have to specify a stochastic mortality rate model for each age since the age naturally plays an important role for a person's likelihood of dying. This suggests to model the time of death τ_{x_0} of an individual of age x_0, as the first jump time of a doubly stochastic point process with intensity $\lambda(x_0 + t, X^1(t))$. in terms of Definition 2.1.1, we therefore consider the following setup:

Definition 2.8.1 *The time of death of an x_0 year old person at inception is modeled as in Definition 2.1.1 with*

$$\Lambda_{x_0}(t) = \int_0^t \lambda_{x_0}(s)\, ds,$$

where we call

$$\lambda_{x_0}(s) := \lambda(x_0 + s, X^1(s))$$

the mortality intensity*.*

As a consequence of Definition 2.8.1, the mortality intensity, which gives the instantaneous probability for dying of an $x_0 + s$ year old person at time s, is influenced

by two sources: First, there is a deterministic component related to the age of the person. Second, mortality intensities are influenced by the background process X^1, which evolves independently of deaths and represents the stochastic nature of future mortality rates; X^1 can symbolize general market factors such as interest rates, too. Furthermore, our specification of the death times rules out simultaneous deaths (see Table 2.1), which is not a problem since the realizations of the jumps itself are not relevant anymore as argued above.

Generally, our definition of the mortality intensity entails the conditional independence setup considered in Proposition 2.5.1. In particular, from Proposition 2.5.1 it follows that expectations contingent on survival of the person can be written as[37]

$$\mathbb{E}\left[1_{\{\tau_{x_0}>T\}}\Upsilon\middle|\mathcal{F}_t\right] = 1_{\{\tau_{x_0}>t\}}\mathbb{E}\left[\exp\left\{-\int_t^T \lambda_{x_0}(s)\,ds\right\}\Upsilon\middle|\mathcal{F}_t^{X^1}\right],$$

and, in particular, the conditional survival probability of the person over $[t,T]$ is given by

$${}_{T,t}p_{x_0} = \mathbb{E}\left[\exp\left\{-\int_t^T \lambda_{x_0}(s)\,ds\right\}\middle|\mathcal{F}_t^{X^1}\right].$$

In the literature on stochastic mortality modeling, the model approach presented here, which is a special case of the general setup of Definition 2.1.1, is often referred to as the *spot force mortality modeling approach* since by specifying a model for $\lambda_{x_0}(s)$ we model the instantaneous rate of dying per unit time, i.e. the so-called *spot force of mortality*.

Within our setup – instead of modeling the spot force of mortality – we could also setup a model of the "forward" intensities similar to the Heath-Jarrow-Morton approach for modeling interest rates, see for example Bauer and Russ (2006). For a general overview and a categorization of continuous mortality models, see Cairns et al. (2006). However, the more "direct outcome" of the general setup considered here are spot force models; hence, most model specifications proposed within our setup in the literature belong to this class.

In particular, the application of affine processes (see Section 2.7) to stochastic mortality modeling has been brought forward by a number of authors due to their relatively nice properties (see e.g. Biffis (2005), Dahl (2004), or Schrager (2006)). These nice properties carry over to the more general setup studied in Section 2.7, which extends the affine modeling approach: Given that $X^1 = (Y^1, Y^2, Z)$ of Definition 2.8.1 satisfies the conditions stated in Assumption 2.7.1 and that $\lambda_{x_0}(s)$ can be written as

$$\lambda_{x_0}(s) = \lambda(x_0+s, X^1(s)) = Z(t)\chi^1(x_0+s) + \chi^2(x_0+s)Y^2(s)$$

[37] Under the technical integrability conditions regarding Υ stated in Proposition 2.2.2.

with $\chi^1(t) \in \mathbb{R}^r_{0+}$ and $\chi^2(t)^T \in \mathbb{R}^{d_2}_{0+}$ positive functions for all $t \geq 0$, we have

$$ {}_{T,t}p_{x_0+t} = e^{Z(t)u(t,T)+v(t,T)Y(t)}. $$

Here, u and v solve equations (2.13) and (2.14) with appropriate parameters but additionally depend on the person's age x_0 at t since χ^1 and χ^2 are functions of $x_0 + t$.

In a concrete specification, χ^1 and χ^2 usually represent parametric functions of age, i.e. mortality laws $f(x_0 + s; \theta)$. An example of such a mortality law, is the so-called Gompertz law:

$$ f(x_0 + s; \theta = (a, b)) = e^{a(x_0+s)+b}. $$

We end this subsection by stating a general result regarding the dependence structure between the survival events of a cohort; it is a corollary of Theorem 2.5.2:

Corollary 2.8.1 *The single survival probabilities ${}_{T,t}p_{x_0}$ of the persons with age x_0 at inception are transformed to the joint survival probability as implied by an Archimedean copula with generator $\phi = {}_{T,t}\varphi^{-1}_{\int_0^\bullet \lambda_{x_0}(s)ds}$, where*

$$ {}_{T,t}\varphi_{\int_0^\bullet \lambda_{x_0}(s)ds}(u) = \mathbb{E}\left[\exp\left\{-u\int_t^T \lambda_{x_0}(s)\,ds\right\}\middle|\, \mathcal{F}_t^{X^1}\right]. $$

2.9 Summary and Remarks

This chapter introduces a flexible framework for modeling a vector of stopping times and builds the foundation for the analysis of structured finance products in Chapters 3 and 4. In the proposed stopping times model, conditional survival probabilities depend on past realizations of the τ_is as well as on some exogenous background process X^1. This process can model factors that influence the jumps of N but do not depend on these jumps. For example, in a credit portfolio application it is usually reasonable to assume that the general macroeconomic environment has an impact on the default likeliness of single firms but not vice versa. Furthermore, realizations of the τ_is can coincide in our setup, i.e. simultaneous jumps of the N_i are possible.

We provide a rigorous mathematical discussion of the introduced framework, demonstrate that it is well-defined and show how important quantities such as survival probabilities can be calculated. In addition, we state conditions under which the setup reduces to other setups which have already been studied in detail in scientific literature such as the Lando (1998) setup or the common Poisson shock models of Lindskog and McNeil (2003). As a further contribution, we analyze the dependence structure between the stopping times as implied by the model from

two different perspectives. The first perspective is eventually a static one meaning that we consider the dependence structure between the survival events over a fixed time horizon. We find that the model-implied dependence structure is the one implied by the copula function studied in Marshall and Olkin (1988). Moreover, we examine the dynamics of the process counting the occurrence of the stopping times and introduce a new measure to describe the volatility of this process. This represents a dynamic approach to characterize dependencies between the stopping times because we measure the clustering of the events in time.

Our findings provide a deeper understanding of the dependence structure implied by many models proposed in the literature since most concrete model specifications can be formulated within our general framework. They further help to structure the vast number of credit portfolio and CDO models which have been published in the recent past. As our results show, some specifications imply in fact similar or even almost identical models. For example, we show that the Joshi and Stacey (2006) model eventually is a common Poisson shock model (cf. Example 2.3.1). Also, we provide a detailed discussion of the Duffie and Gârleanu (2001) model based on the developed theoretical tools.

Moreover, we investigate in detail the question of how to arrive at models that guarantee a certain degree of analytical tractability within our setup. In particular, we identify situations where the distribution of the loss process as well as survival probabilities can be calculated at least semi-analytically. Finally, we establish a link between our general setup and the more specific setups considered for the applications in Chapters 3 and 4.

Future research should contain a further analysis of the aggregate losses implied by different model specifications. For example, Cousin and Laurent (2007) consider the ordering of stop-loss premiums and convex risk measures on aggregate losses for the case of exchangeable Bernoulli-mixture models which are a special case of the conditional independence setup encountered in this chapter. An interesting question would be, how results change when allowing for contagion effects and how contagion models can be compared with conditional independence models.

Chapter 3

Application I: Modeling of Structured Credit Products

Corporate defaults tend to cluster in time and in industries. In 2002, for example, the annual corporate default rate recorded by Moody's Investors Service (2003) was 3%, more than twice the long-run average; telecommunications issuers like Global-Crossing or Worldcom constituted 31% of all defaulted issuers. While the clustering of defaults is already important for the management of credit portfolios, it becomes even more important when analyzing structured credit products. A CDO tranche, for example, is a contingent claim on a debt portfolio. Cash flows from the debt portfolio are first used to serve the most senior tranche. Remaining funds are then distributed according to the tranches' rank in the seniority ladder. By construction, the single tranches reference different "slices" of the underlying portfolio's loss distribution. For instance, the risk profiles of the more senior tranches relate to the tail of the loss distribution. Even one of the tranche's simplest risk measures, its expected loss, strongly depends on the portfolio dependence structure. By investing into tranches, investors can therefore take a position in the dependence structure of the underlying portfolio, and the trading with structured credit products can be considered as a "dependence structure" trading.[1] Consequently, the key challenge, which we face when modeling structured credit products, is to correctly model the underlying portfolio and, in particular, its dependence structure. Missspecified dependence structures will affect the forecasted loss distribution of the structured credit product in a non-linear way.

In standard credit risk models, default clusters are usually explained by common factors like the business cycle or stock market valuations that induce correlations in individual firms' default probabilities. Conditional on the stochastic evolution of common factors, defaults are independent (cf. Subsection 2.5.1). The approach is ubiquitous in financial institutions. Key applications range from industry mod-

[1] Not surprisingly, the financial industry often speaks of "correlation" trading.

els of portfolio credit risk (see e.g. Crouhy et al. (2000)) and the new capital requirements (e.g. Gordy (2003)) to the pricing of structured finance instruments like collateralized debt obligations (e.g. Duffie and Gârleanu (2001)).

Up to date, most of the time-continuous models discussed in scientific literature have been studied in a pricing context (see e.g. Mortensen (2006), Feldhütter (2008) or Kiesel and Scherer (2007)). However, less is known about the performance of these models under the objective, i.e. the real world, measure. The present chapter aims to close this gap. Within our general stopping times model introduced in the previous chapter, we provide a detailed empirical investigation of different model specifications in the context of a data set that consists of 250090 default intensities for 3241 different firms. Considered models comprise purely diffusion-based models, specifications with jumps and models which are additionally driven by regime processes. Also, the importance of contagion effects for explaining defaults is investigated, but since we find these effects of minor importance we primarily work with models that satisfy the conditional independence assumption.

The remainder of this chapter is structured as follows: Section 3.1 contains the estimation of the firms' default intensities and investigates whether the intensities are able to explain the observed default behavior. In Section 3.2, we introduce a model for the joint intensity dynamics and discuss its properties. The calibration of the model on a single firm basis follows in Section 3.3. In Section 3.4, we consider the issue of model risk in connection with the conditional independence assumption. Section 3.5 provides the estimation of the portfolio model and derives transition matrices of structured credit products. Section 3.6 concludes.

3.1 Estimation of the Firm's Default Intensities

In this section, first default intensities for a large number of US and non-US corporates are estimated. In a second step, we then assess the quality of our estimations with respect to default prediction and default clustering.

For the estimation, we use a model that satisfies the conditional independence (or equivalently doubly stochastic) assumption (cf. Subsection 2.5.1). More precisely, we presume that defaults are independent conditional on observable factor variables and that these factor variables themselves are not affected by defaults in the portfolio. Das et al. (2007) conducted the first rigid empirical test of this assumption. They test whether defaults of US corporates are consistent with the model-implied default clustering of the intensity model introduced by Duffie et al. (2007). Das et al. (2007) find that observed default clustering exceeds the one implied by the estimated model. Therefore, they conclude that the doubly stochastic assumption

is invalid and that credit risk models should be enriched by contagion effects or "frailty", i.e. unobservable variables, in order to account for the extra correlation. Otherwise, one might incur significant errors in the assessment of credit portfolio risk or the pricing of structured finance products.

As pointed out by Das et al. (2007), their statistical tests are joint tests of well-specified default intensities and the doubly stochastic assumption.[2] In our tests, we use a data set that is very similar to the one used by Das et al. (2007) but introduce two modifications to the estimation of default intensities. First, we model intra-month patterns in observed defaults. Second, we estimate default intensities on an out-of-sample basis, which brings our estimates closer to the ones financial institutions implementing the models would actually have used. We find that both modifications increase the ability of the intensity model to explain observed default clustering. When both modifications are made, the test suggested by Das et al. (2007) no longer rejects the validity of the doubly stochastic assumption.

The remainder of the section is organized as follows. After introducing the general model in Subsection 3.1.1, we describe our data set in Subsection 3.1.2. Our specification of the default intensities and their estimation follows in Subsection 3.1.3; the predictive power of the model is assessed in this subsection, too. Subsection 3.1.4 analyzes whether the estimated models are able to account for the default clustering observed in the data.

3.1.1 Model

As previously mentioned, for the estimation of the default intensities we rely on a conditional independence model.[3] More precisely, in terms of Definition 2.1.1 we assume that

$$\Lambda_i(t) = \int_0^t \lambda_i(s, X^1(s))\, ds.$$

Since the σ-fields $\bigvee_{1\le i\le I} \sigma(E_i)$ and $\mathcal{F}_{T^*}^{X^1}$ are independent, the defaults of the portfolio objects are conditionally independent with respect to $\mathbf{F}^{X^1}$ (cf. Proposition 2.5.1), and simultaneous defaults are ruled out (see Table 2.1). In addition, $\lambda_i(t)\Delta$

[2] As pointed out by Lando and Nielsen (2008) the statistical tests are unfit to uncover a special type of contagion in which the default of one firm affects the factor variables and thus the default intensity of others. A passed test has therefore to be interpreted as follows: Observed defaults are likely to have been generated by the aggregated path of estimated default intensities and one needs not introduce contagion effects *in addition* to the already used explanatory variables to explain the observed default clustering.

[3] This assumption is not significant to the estimation and could be relaxed towards intensities that simply depend on observable factor variables no matter whether these factor variables evolve independently of defaults or not.

presents the instantaneous probability of a firm to default over the next infinitesimally small time step Δ because we have

$$\lim_{\Delta \to 0} \frac{\mathbb{P}\left(t < \tau_i \leq t + \Delta \middle| \, \tau_i > t\right)}{\Delta} = \lambda_i(t).$$

In the concrete model implementation, which we consider in Subsection 3.1.3, the background process $X^1 = (R^1, R^2)$ will be given by the explanatory variables or *covariates*, where in this chapter

- R^1 will represent macroeconomic variables such as the S&P 500 index return, and
- $R^2 = (R_1^2, \ldots, R_I^2)^T$ will summarize all firm-specific variables such as the firms' long-term debt rating.

As a consequence, the dimension of R^1 simply corresponds to the number of macroeconomic variables used, while the dimension of R^2 is the number of firms in the portfolio times the number of firm-specific variables considered.

Furthermore, as is common in the literature we will consider a parametric function for the default intensities λ_i such that

$$\lambda_i(t, X^1(t)) := \lambda\left(t, \theta, R^1(t), R_i^2(t)\right)$$

with $\theta \in \mathbb{R}^d$ some parameter vector. Given θ and a path $x^1 = (r^1, r^2)$ of X^1, the likelihood $\mathcal{L}(\theta; r^1, r^2)$ of observing some vector $\tau = (\tau_1, \ldots, \tau_I)$ of default times is given by

$$\mathcal{L}(\theta; r^1, r^2) = \prod\nolimits_{i \in I} e^{-\int_0^{\tau_i} \lambda\left(u, \theta, r^1(u), r_i^2(u)\right) du} \lambda\left(\tau_i, \theta, r^1(\tau_i), r_i^2(\tau_i)\right), \tag{3.1}$$

since the probability of firm i to default at $t = \tau_i$ conditional on the path of X^1 is given by $e^{-\int_0^t \lambda\left(u, \theta, r^1(u), r_i^2(u)\right) du} \lambda\left(t, \theta, r^1(t), r_i^2(t)\right)$, and conditional on a path of the background process X^1 the default times are independent (cf. Proposition 2.5.1). To obtain the conditional density of τ_i, we simply differentiate the conditional survival probability.

For the portfolio loss process,

$$L(t) := \sum_{i=1}^{I} 1_{\tau_i \leq t},$$

it follows that

$$M(t) := L(f^{-1}(t))$$

is a standard Poisson process (cf. Theorem 2.6.1) where
$f(s) := \sum_{i=1}^{I} \int_0^s 1_{u<\tau_i} \lambda_i(u, X^1(u))du$, i.e. its increments

$$M(T) - M(t) \sim Poi(T - t). \tag{3.2}$$

In addition, we have (cf. Corollary 2.6.1)

$$\sum_{i=1}^{I} \int_{\tau_{(i-1)}}^{\tau_{(i)}} 1_{u<\tau_i} \lambda_i(u, X^1(u))du \sim Exp(1), \tag{3.3}$$

where $\tau_{(0)} := 0 < \tau_{(1)} < \ldots < \tau_{(I)}$ denote the ordered default times of the portfolio objects. Inspired by Das et al. (2007), we will make use of this relationship in Subsection 3.1.4 when testing whether the estimated models are able to explain the observed default clustering.

3.1.2 Data

We use data on US and non-US corporates which have a traded equity and a Moody's rating. The data extend from January 1980 to April 2005. The information provided include daily default times and two firm-specific variables that we are going to use as default predictors:

- The firms' one-year *Expected Default Frequencies* (EDFs) provided by Moody'sKMV, which are month-end values ranging from 0.02% to 20%; a firm's one-year EDF is a non-parametric estimator of its one-year default probability based on the firm's "distance-to-default" which itself is a leverage measure adjusted for asset volatility that goes back to the firm value model of Merton (1974). For a more detailed description of the EDF as a measure of a firm's default risk see for example Berndt et al. (2005) who use the EDF in order to estimate and analyze risk premia of corporate bonds.

- The Moody's long-term rating of the firms with values {"Aaa", "Aa1", ..., "C"} which we transform to the cardinal numbers $\{1, \ldots, 21\}$ where 1 corresponds to a "Aaa" Rating and 21 to a "C" rating. Rating actions such as downgrades or upgrades are reported on a daily basis.

Our use of both EDFs and ratings as explanatory variables for default prediction is inspired by Löffler (2007) who shows within a static logit regression model that adding ratings to EDFs increases predictive power.

The original data set contains observations for which one of the two variables (EDF, rating) is missing. A possible solution to this problem would be to extrapolate or interpolate missing EDFs or ratings. However, this could make later results sensitive to the chosen interpolation method. Therefore, we remove data points with

missing variables. Furthermore, we disregard multiple defaults of firms which can be observed in the data set, i.e. observations after the first default of a firm are not taken into account. We also exclude observations of firms that are a 100% subsidiary of a holding and that defaulted together with this holding at the same date.[4]

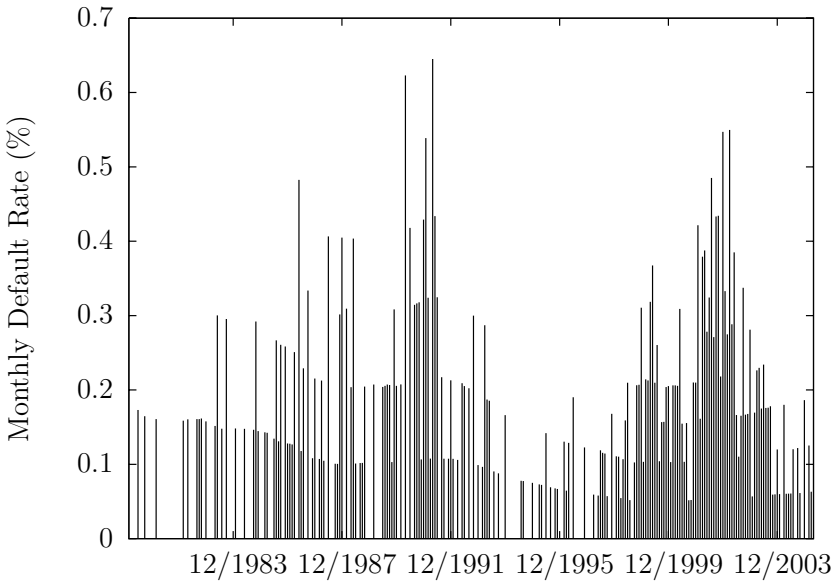


Figure 3.1: Monthly default rate in the data set in the time period from 02/1980 to 04/2005.

Finally, we end up with a data set consisting of 370345 observation months for 3989 different firms across the time period from 01/1980 to 04/2005. The sample includes 511 defaults which is more than the 495 defaults in Das et al. (2007) and the 370 defaults in Lando and Nielsen (2008). The number of firms in the data set increases from 1980 on, reaches its maximum of approximately 1950 firms around the year 1999 and slowly declines afterwards. Figure 3.1 shows the monthly default rates. The recessions in 1990-91 and 2001 are visible with monthly default rates peaking at around 0.6%.

Apart from the stated firm-specific variables we also use macroeconomic variables for the estimation of the intensities. Duffie et al. (2007) point out that one can

[4]More precisely, we removed *Safety-Kleen Corporation* and *Laidlaw One, Inc.* which defaulted together with *Laidlaw Inc.* on 05/15/2000 and *DecisionOne Corporation* that defaulted together with *DecisionOne Holdings Corp.* on 08/02/1999.

gain predictive power by including macroeconomic variables into the estimation. The macroeconomic variables (all monthly) that we use are[5]

- The one-year trailing return on the S&P 500 index.
- The 3-month US-Treasury bill rate.
- The one-year percentage change in US Industrial production, calculated from the gross value of final products and nonindustrial supplies (seasonally adjusted). We use this variable with a 1-month lag since this data is always published in the middle of the next month.
- Spread between the ten-year and one-year treasury rate.

The first two variables have been used in Duffie et al. (2007) and Das et al. (2007) and the last two have been suggested by Lando and Nielsen (2008).

3.1.3 Estimation Results

After having decided on the covariates that influence the default intensities λ_i, it remains to specify a parametric function. For our specification, we rely on a *Proportional Hazards Rate Model*, see e.g. Therneau and Grambsch (2001), in which a firm's default intensity $\lambda_i(t)$ is linked to the observable variables $(R^1(t), R_i^2(t))$ via the relationship

$$\lambda_i(t) = \lambda\left(t, \theta, R^1(t), R_i^2(t)\right) = \lambda^0(t) \exp\left(\sum_{n=1}^{N} \theta_n R_n^1(t) + \sum_{m=1}^{M} \theta_{N+m} R_{mi}^2(t)\right).$$

As already mentioned, $R^1(t) = (R_1^1(t), \ldots, R_N^2(t))$ denotes the time t values of the macroeconomic variables, $R_i^2(t) = (R_{1i}^2(t), \ldots, R_{Mi}^2(t))$ the values of the firm-specific variables of firm i and $(\theta_1, \ldots, \theta_{N+M})$ the coefficients describing how the variables influence the intensity; $\lambda^0(t)$ denotes a possibly time-dependent intercept that may capture recurring intra-month patterns in default times and which we refer to as the *Baseline Component* in the following. So far, our model is the same as utilized by Duffie et al. (2007) and Lando and Nielsen (2008) except for the choice of firm-specific variables and for the Baseline Component. Differences in firm-specific variables should be minor, though. Both Duffie et al. (2007) and Lando and Nielsen (2008) employ a measure of distance-to-default which is also at the heart of the EDF that we use.

[5] All macroeconomic variables have been downloaded from the Federal Reserve Board Website.

The Baseline Component

Since later model specification tests strongly depend on the clustering of defaults over time – even on the clustering within small time intervals such as months – the estimation of the firms' default intensities described next will take the timing of defaults within the corresponding default months into account, too. Figure 3.2 shows how defaults are distributed across the single days of a month. Defaults are heavily clustered around the 1st and the 15th of each month, presumably because default is often declared after a missed coupon payment (the 1st and the 15th are common choices for coupon payment dates). Loosely speaking, if a firm has survived the 15th of a month, it has very good chances to survive the whole month since the dates around the 1st and the 15th account for more than 60% of the defaults.

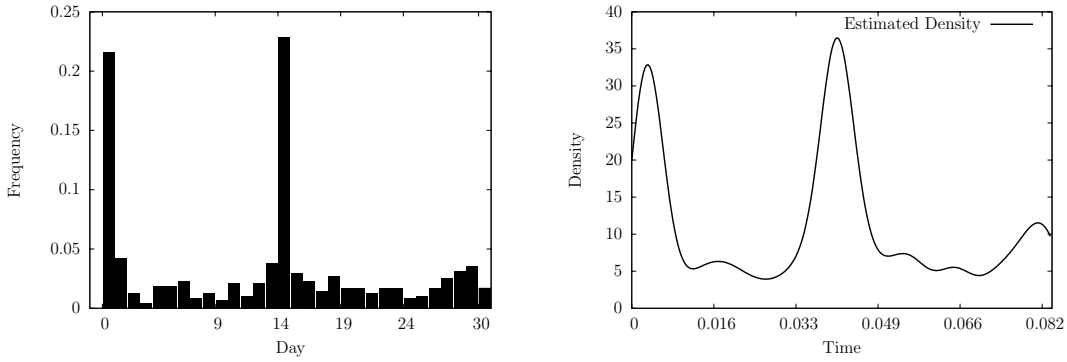


Figure 3.2: Histogram of intra-month default dates in the data set (left) and corresponding estimated density (right).

As baseline hazard rate we therefore use a periodical function with period 1 month of the form

$$\lambda^0(t) := \frac{1}{12}\, e^{\theta_0}\, h\left(t - \frac{[12 \cdot t]}{12}\right),$$

where $[\cdot]$ rounds any real argument to the next lower integer and h represents the kernel density estimator associated with the observed default frequency on the different dates; this kernel density estimator is plotted in Figure 3.2, too. To estimate the density, we used 256 nodes and a Gaussian kernel with appropriate bandwidth. The calculated Baseline Component is used in all estimations even if we apply the model out-of-sample since we assume that the intra-month default timing is fixed over time. To support this assumption, as a robustness check, we calculated the densities of the first 50% and the last 50% of defaults in the data set. These are presented in Figure 3.3. The distribution seems to remain stable over

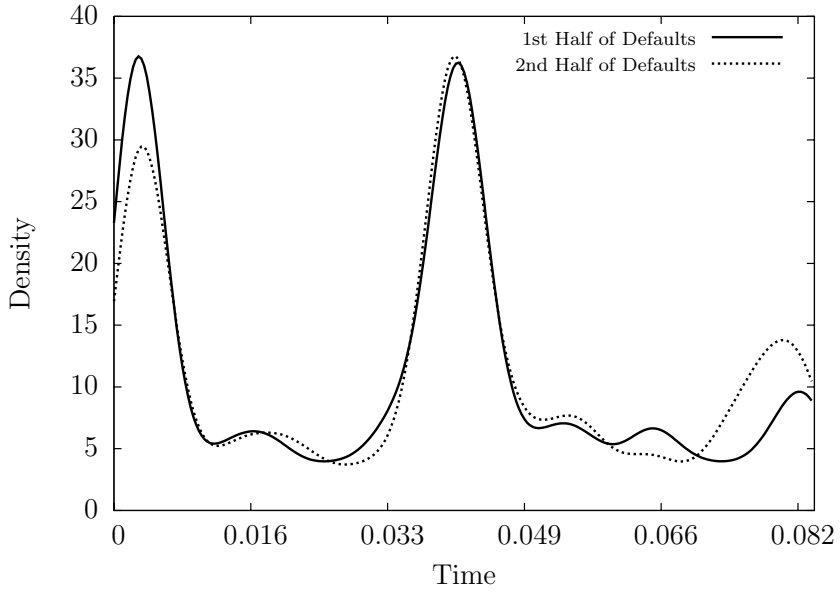


Figure 3.3: Estimated density of intra-month default dates based on the first or the second half of defaults in the data set.

time. Furthermore, since $\int_t^{t+\frac{1}{12}} \lambda^0(s)ds = \frac{1}{12}e^{\theta_0}$, the Baseline Component leaves a firm's default probability over the next month unchanged and does only very weakly change the estimated values of the coefficients $(\theta_0, \ldots, \theta_{N+M})$, i.e. estimated coefficients $(\hat{\theta}_0, \ldots, \hat{\theta}_{N+M})$ almost coincide for a model with and without Baseline Component.

Estimation Results with Fixed Coefficients

Having specified the Baseline Component, we have finally set the stage for conducting model inference via ML. We start by estimating intensities in-sample, i.e. coefficients are fixed for the entire sample period. Since observations for individual firms may start at different dates, the likelihood function (3.1) that is actually used in the estimation has the form

$$\begin{aligned}\mathcal{L}(\theta; r^1, r^2) &= \prod_{i \in I} \exp\left(- \int 1_{u \in \mathcal{T}_i, u < \tau_i} \lambda(u, \theta, r^1(u), r_i^2(u)) du \right) \\ &\qquad \cdot \left(1_{\tau_i \in \mathcal{T}_i} \lambda(\tau_i, \theta, r^1(\tau_i), r_i^2(\tau_i)) + 1_{\tau_i \notin \mathcal{T}_i} \right),\end{aligned}$$

where $\mathcal{T}_i$ denotes the union of intervals for which all variables of firm i that we want to use (default information, rating, EDF) are available. Since we disregard multiple defaults, $\max \mathcal{T}_i = \tau_i$ in case of a default of i.

The ML estimator $\hat{\theta}$ of θ is then given as

$$\hat{\theta} = \arg\max_{\theta} \ \mathcal{L}(\theta; r^1, r^2).$$

Although rating information is available on a daily basis, we only use the month-end rating implying that observable variables are constant during a month and changes in the intensity during that time are only due to variations of $\lambda^0(t)$. Ratings enter in the coding from 1 (Aaa) to 21 (C) described above; the EDF enters as logarithm of EDF. As documented in Löffler (2007), these simple transformations work relatively well compared to other alternatives.

	Models			
	I	II	III	IV
Constant	-7.08	-13.2	-7.45	-14.8
	(29.2)	(31.6)	(24.4)	(28.8)
S&P 500	-0.0210	-0.199	0.600	0.438
	(0.0748)	(0.704)	(1.94)	(1.43)
3-month rate	-0.00901	0.152	0.0784	0.306
	(0.357)	(5.71)	(2.39)	(8.99)
Industrial Production	–	–	-0.0787	-0.0588
	–	–	(4.20)	(3.05)
Spread 10year-1year	–	–	0.153	0.362
	–	–	(2.40)	(5.67)
Log(EDF)	2.08	1.40	2.08	1.40
	(26.4)	(17.0)	(26.9)	(17.0)
Rating	–	0.448	–	0.475
	–	(18.9)	–	(19.0)
Log-Likelihood	231	425	245	449

Table 3.1: ML estimates of intensity models in four specifications (in-sample). Each model includes the estimated Baseline Component. T-statistics in parentheses.

The estimated default intensity parameters $\hat{\theta}$ as well as their t-statistics are displayed in Table 3.1 for the basic four model specifications which we will consider throughout the text.[6] Here as in all later tests, we restrict the sample period from 01/1985 to 04/2005 because this is the time period for which we can estimate the intensities with our alternative approach (rolling estimations with a five-year estimation window). Using the entire available data for the fixed coefficients estimation does not lead to qualitatively different results.

[6] t-statistics are derived by inverting the Fisher-Information matrix at $\theta = \hat{\theta}$.

In most cases the coefficients exhibit the expected sign. Default intensities are increasing in EDF, rating and they are decreasing in industrial production.[7] Exceptions are the S&P 500 index, the spread between the 10 year and the 1 year rate and the 3-month US-Treasury bill rate. At first sight, one would probably have expected all three coefficients to have a negative sign because boom periods (with defaults below average) are usually accompanied by positive stock returns, high interest rates and an upward sloping yield curve. However, the sign of the S&P 500 index is only positive *after* controlling for the other variables. An increased survival probability of firms during a stock market boom might already be captured by variation in the EDF. Short-term interest rates could go along with higher default rates because they impact borrowing costs or mark the end of a boom. Note that in Duffie et al. (2007) the coefficients on the S&P 500 is positive, too; the one on short-term interest rates is negative. The coefficient for the spread between the 10 year and the 1 year rate is positive in Lando and Nielsen (2008), too.

Estimation Results with Rolling Estimation Windows

The in-sample estimation from the previous section presumes that the functional relationship between explanatory variables such as the EDF or rating and the intensities is constant over a time period of 25 years. Even though default information providers like Moody's or Moody'sKMV aim at the time-consistency of their default risk assessments, it is unclear whether they achieve perfect consistency. Two possible reasons for non-stationary behavior are changes in the information content of accounting measures (cf. Jorion et al. (2008)) or sampling error in the estimation and calibration of risk measures. EDFs, for example, use estimated stock market volatilities and are calibrated on past default behavior which exposes them to sampling error; sampling errors can be correlated across firms and therefore affect the average precision of EDFs. In addition, the information content of macroeconomic variables such as the term spread can change over time (see e.g. Benati and Goodhart (2007)).

Therefore, working with constant coefficients over long time periods is possibly inappropriate. In order to account for possible variations of the covariates' influence on default intensities, we allow the coefficients to change over time. More precisely, for each $\theta(t)$ we determine an estimation period $\mathcal{U}(t)$ which ends in t and estimate $\theta(t)$ based on the observed covariates and defaults during this period. The forecasts of the intensities are therefore obtained on an out-of-sample basis, which provides another motivation for favoring this approach. We only use information that a financial institution which implemented the model would have had at the time of

[7]Note that a higher rating is a worse rating due to the chosen transformation.

implementation. When we later test whether observed default clustering is consistent with the doubly stochastic approach, we therefore can interpret the results in the sense that actual users of that approach would have incurred significant errors in predicting default clusters or not. With in-sample estimates, test results could be biased in some unknown way.

With the estimation sample restricted to $\mathcal{U}(t)$, the likelihood function of $\theta(t)$ has the form

$$\begin{aligned}\mathcal{L}(\theta;\mathcal{U}(t),r^1,r^2) &= \prod_{i\in I}\exp\left(-\int 1_{u\in\mathcal{T}_i\cap\mathcal{U}(t)}\lambda(u,\theta,r^1(u),r_i^2(u))du\right)\\ &\quad\cdot\left(1_{\tau_i\in\mathcal{T}_i\cap\mathcal{U}(t)}\lambda(\tau_i,\theta,r^1(\tau_i),r_i^2(\tau_i))+1_{\tau_i\notin\mathcal{T}_i\cap\mathcal{U}(t)}\right)\end{aligned}$$

and

$$\hat{\theta}(t)=\arg\max_{\theta}\ \mathcal{L}(\theta;\mathcal{U}(t),r^1,r^2).$$

Our proceeding eventually yields a time series of $\hat{\theta}(t)$.

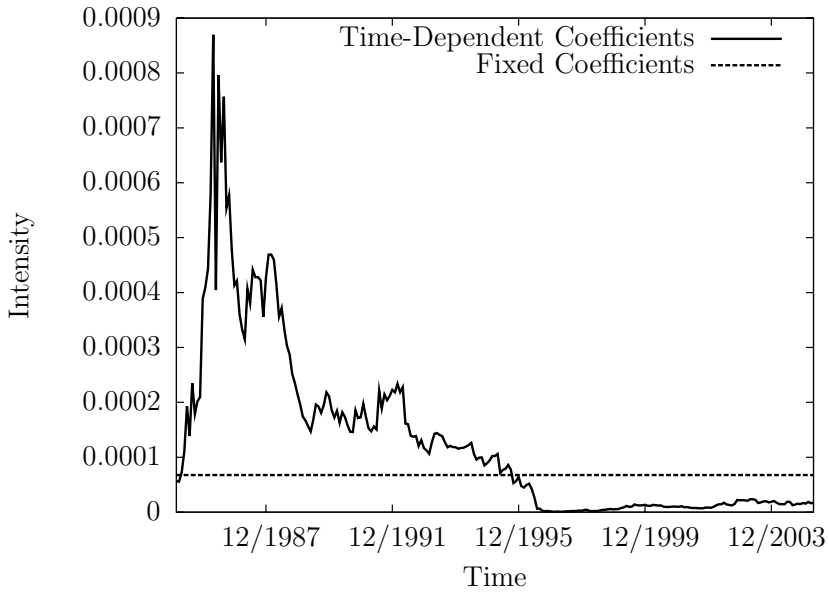


Figure 3.4: The default intensity of a hypothetical firm whose covariates are constant over time based on rolling estimations with five-year estimation windows.

Figure 3.4 illustrates the effects of changing coefficients on intensity estimates. We fix both firm-specific and macroeconomic variables at their means and use the estimated coefficients $(\hat{\theta}_0(t),\ldots,\hat{\theta}_{N+M}(t))$ to determine the intensity in t. Thus, changes in the intensity solely originate from changes in the coefficient estimates.

Because of the five-year estimation period used, the time series $\hat{\theta}$ starts with January 1985 and each of its values $\hat{\theta}(t)$ reflects the past default behavior across the last 5 years. For example, the value of the coefficients at the end of January 1985 are based on the relation between defaults and observable variables during 02/1980 to 01/1985.

The intensity which is implied by the corresponding estimation based on fixed coefficients is also displayed in Figure 3.4. We find that until December 1995 the time-varying intensity lies above the fixed one whereas after 1995 it drops below. This indicates that around 1995 the meaning of some of the variables used in the estimation might have changed.

We choose the five-year estimation period because it is a common choice in the financial industry and because it appears to achieve a good balance between opposing effects. On the one hand, increasing the estimation period should lower estimation errors. On the other hand, it could mean that we use more outdated historical data. Also, it would force us to ignore more observations when testing the models' performance because the first estimates are only available after the difference between the current date t and the start of the data is larger than the estimation period.

Predictive Power

While the focus of this section is on the validity of the doubly stochastic assumption, the test we will use is a joint test of the doubly stochastic assumption and well-specified intensities. It is therefore worthwhile to investigate the models' predictive ability. We follow the literature and use the *Accuracy Ratio AR* to measure the ability to rank firms according to their default probability.

The accuracy ratio is based on the so-called *power curve*. The power curve $pc : [0,1] \mapsto [0,1]$ associated with a model is a function that can be obtained by ranking all firms of the portfolio according to some criterion, which is supposed to carry as much information as possible on the firms' default probability over the period one is interested in. Usually, this criterion would be the firms' default probability over this time period implied by the estimated model. The power curve pc maps the fraction x of worst ranked firms onto the fraction of defaults which these firm account for; it tells us that when we pick the worst ranked firms, a fraction x of all firms accounts for $y = pc(x)$ of the defaults.

We follow Duffie et al. (2007) and define the accuracy ratio to be twice the area between the power curve and the 45° line, i.e.

$$AR = 2 \int_0^1 (pc(x) - x)\, dx.$$

The 45° line is the power curve one would expect to obtain by ranking firms completely randomly. By definition, the maximum value of the accuracy ratio is 1 minus the fraction of actually defaulted firms.

It remains to choose a criterion for ranking the firms. A natural candidate would be the default probability over the prediction period, e.g. the one-year default probability if the prediction horizon is one year. However, to derive such a probability, we would have to set up a model for the dynamics of the firms' default intensities. Therefore, instead of horizon-matched default probabilities we simply use the firms' current default intensities in order to rank them.

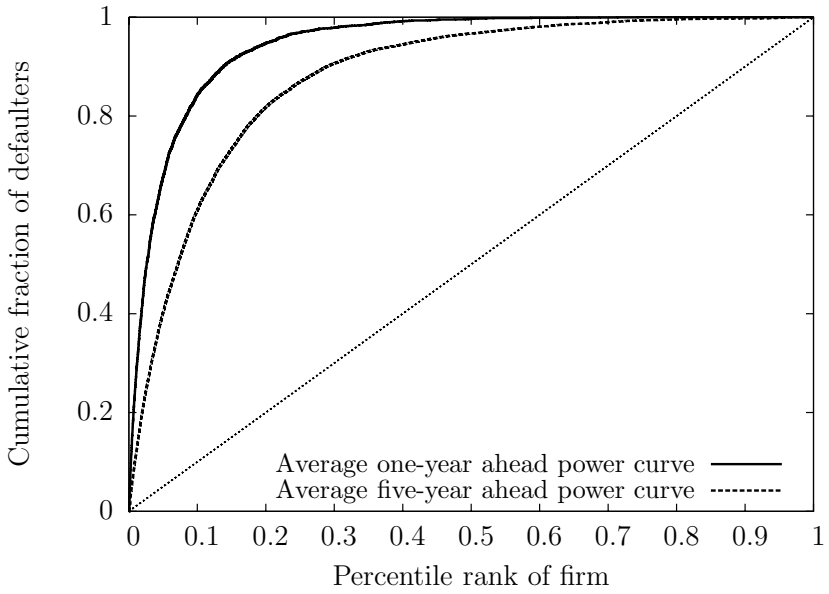


Figure 3.5: The 45° line (dotted) and average one-year and five-year ahead power curves, averaged from 1985 on. Intensities are estimated with model IV and rolling five-year estimation windows.

In Figure 3.5 averaged one-year and five-year ahead power curves are displayed. As awaited, the one-year ahead power curve lies above the five-year ahead power curve since the predictive power of the firms' current default intensities is expected to decrease with increasing forecast horizons. Note that a firm's current default intensity only provides its default probability over the next small time step.

Figure 3.6 shows the time series of one-year and five-year ahead accuracy ratios from 1985 for a model with rating (model IV) and a model without rating (model

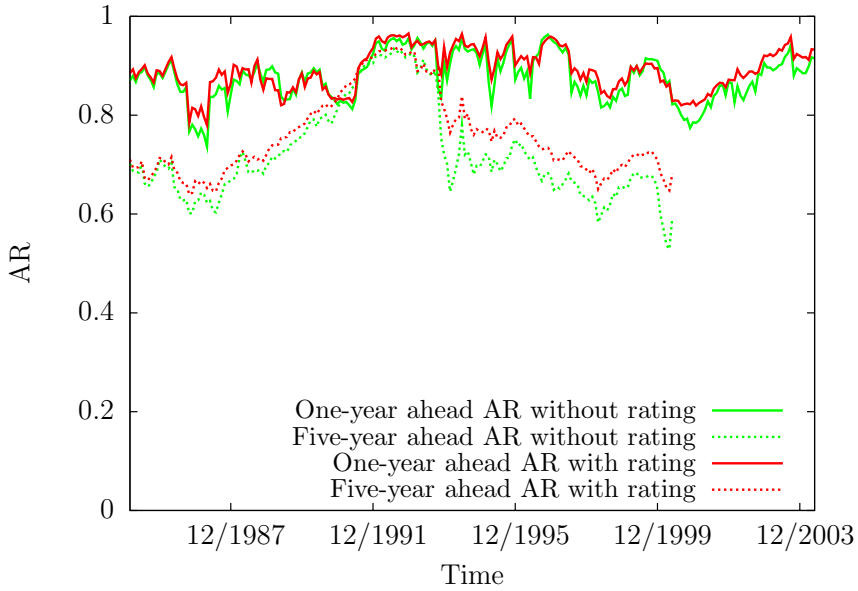


Figure 3.6: One-year and five-year ahead accuracy ratios for a model with rating (model IV) and for a model without rating (model III) from 1985 on. Intensities are estimated with rolling five-year estimation windows.

III). From 1992 on the higher predictive power of a model with rating is visible. Note that the time series of five-year ahead accuracy ratios ends in 2000 due to the chosen five-year prediction horizon, whereas the time series of one-year ahead ratios ends in 2003.

Finally, Table 3.2 summarizes results for default prediction horizons of one and five years, respectively. It shows accuracy ratios for the model with fixed coefficients, which are therefore in-sample, as well as accuracy ratios for the rolling estimation. The accuracy ratios of the model with fixed coefficients are higher, which is unsurprising due to the in-sample nature of these values, but the difference is quite small. Furthermore, in case that the firm rating is not used as a covariate, the obtained accuracy ratios are about the same as the ones reported by Duffie et al. (2007) for a comparable data set over the time period from 1993 on. After taking the firm rating into account, these accuracy ratios considerably increase showing that if one wants to sort firms according to their default probability, the firm rating should be taken into account in the intensity estimation.

We do not want to conceal a problem with the definition of the accuracy ratio used in this investigation: Low default rates can favor high accuracy ratios since their maximum value depends on the fraction of defaulters. We follow this definition

	1985-2005		1993-2005	
	1-year horizon	5-year horizon	1-year horizon	5-year horizon
Fixed coefficients (in-sample)				
Model I	0.878	0.723	0.883	0.697
Model II	0.893	0.765	0.901	0.760
Model III	0.878	0.723	0.883	0.697
Model IV	0.893	0.765	0.901	0.761
Rolling estimation (out-of-sample)				
Model I	0.878	0.723	0.883	0.697
Model II	0.891	0.759	0.900	0.751
Model III	0.878	0.723	0.883	0.697
Model IV	0.891	0.760	0.900	0.752
Compared to prior studies (out-of-sample)				
Duffie et al. (2007)			0.88	0.69

Table 3.2: Accuracy ratios of the different regression intensity models averaged from 1985 on and from 1993 on. Rolling estimations have been conducted with five-year estimation windows.

in order to keep our results comparable, because it has been also used in a prior study by Duffie et al. (2007). To make results even more comparable, however, it would be desirable that all studies use an accuracy ratio which is normalized by the fraction of firms that have survived (for such a definition see Tasche (2007)).

3.1.4 Comparing Model-implied and Observed Default Behavior

In this subsection, we assess the quality of our intensity estimations with respect to model-implied

- default clustering and
- single default probabilities.

In particular we test whether observed defaults are likely to have been generated by estimated intensities. For investigating the default clustering, on the one hand, we will consider defaults and intensities on an aggregate, i.e. portfolio, basis. On the other hand, when testing the estimated models for well-specified single default probabilities we examine defaults and default intensities on a single firm basis.

Default Clustering

We now assess whether the intensity estimates provide default forecasts which are consistent with observed default clustering. For this purpose, we will consider defaults and default intensities on an aggregate, i.e. portfolio, basis. More precisely, we explore the question of whether the path of observed portfolio defaults is likely to have been generated by the (aggregated) paths of default intensities that have been estimated in the previous section.

Our analysis of aggregated intensities and defaults is based on a so-called *Fisher's dispersion test*. Das et al. (2007) have introduced this test to the credit risk literature in order to test the doubly stochastic assumption. In the test, one exploits the fact that the time-changed portfolio loss process M is a standard Poisson process (cf. equation (3.2)), i.e.

$$M(t+\delta) - M(t) \sim Poi(\delta).$$

By dividing the total estimated aggregated intensity $\sum_{i=1}^{I} \int_0^\infty 1_{u<\tau_i} \lambda_i(u) du$ into bins of equal size δ, we obtain a series of i.i.d. $Poi(\delta)$-distributed random variables. More precisely, we calculate calendar times $t_0, t_1, t_2, \ldots$ such that $\sum_{i=1}^{I} \int_{t_i}^{t_{i+1}} 1_{u<\tau_i} \lambda_i(u) du = \delta$. Then, the

$$U_k := M(k\delta) - M((k-1)\delta) = \sum_{i=1}^{I} 1_{t_{k-1}<\tau_i \leq t_k}$$

form a series of i.i.d. $Poi(\delta)$-distributed random variables. A simple test if the U_k, $k \in \mathbb{N}$ are indeed Poisson distributed is Fisher's dispersion test (cf. Cochran (1954)): Given $U_k \sim Poi(\delta)$, $k \in \mathbb{N}$,

$$\sum_{k=1}^{K} \frac{(U_k - \delta)^2}{\delta} \sim \chi^2_{K-1},$$

for arbitrary $\delta > 0$.

In Table 3.3, the p-values of Fisher's dispersion test with respect to different bin sizes δ for fixed coefficients as well as for rolling estimations (with 5 year estimation windows) are presented. The tests are performed for 490 defaults over the time period from January 1985 to April 2005 since for this period both estimates are available.

i) Fixed coefficients (in-sample):

Without Baseline Component

	1	2	4	6	8	10	16
Model I (closest to Das et al. (2007))	3.74e-06***	0.00238**	0.0134*	0.0304*	5.86e-04***	0.00497**	8.01e-05***
Model II	7.98e-05***	0.00382**	1.71e-05***	5.96e-08***	2.10e-08***	1.21e-08***	3.45e-09***
Model III (closest to Lando and Nielsen (2008))	0.0352*	0.278	0.630	0.146	0.882	0.302	0.148
Model IV	0.00113**	0.0974	0.0261*	0.0151*	0.00385**	0.00245**	0.0145*

With Baseline Component

	1	2	4	6	8	10	16
Model I	0.0143*	0.0564	0.0182*	0.0515	0.00163**	0.00976**	1.13e-04***
Model II	5.25e-04***	2.64e-04***	7.59e-05***	1.06e-06***	3.53e-08***	1.46e-08***	3.61e-09***
Model III	0.228	0.578	0.726	0.290	0.793	0.386	0.0913
Model IV	0.108	0.349	0.0261*	0.0282*	0.00330**	0.00729**	0.0151*

ii) Rolling estimation (out-of-sample):

Without Baseline Component

	1	2	4	6	8	10	16
Model I	0.157	0.267	0.0955	0.244	0.0229*	0.00206**	7.71e-04***
Model II	0.225	0.671	0.248	0.132	0.138	0.0254*	0.00178**
Model III	0.0214*	0.268	0.313	0.0660	0.117	0.0321*	0.0569
Model IV	0.0681	0.329	0.223	0.221	0.0795	0.0302*	0.0740

With Baseline Component

	1	2	4	6	8	10	16
Model I	0.340	0.540	0.0909	0.189	0.0318*	6.83e-04***	7.71e-04***
Model II	0.788	0.908	0.580	0.123	0.195	0.0292*	0.00215**
Model III	0.145	0.296	0.282	0.0988	0.106	0.0180*	0.133
Model IV	0.930	0.955	0.520	0.418	0.0880	0.0359*	0.0910

Table 3.3: p-values of Fisher's dispersion test for bin sizes 1, 2, 4, 6, 8, 10, 16. Intensity models I to IV differ in explanatory variables (models III and IV add the term spread and industrial production, models II and IV add the rating). Results are presented for all combinations of two further estimation specifications: in-sample or out-of-sample estimation, estimation with or without Baseline Component. The Baseline Component captures the intra-month pattern of defaults. Low p-values would lead to rejection of the doubly stochastic assumption.

With fixed coefficients, no Baseline Component and the macroeconomic variables from Das et al. (2007), the levels of the p-values are similar to the ones derived by Das et al. (2007) for the Duffie et al. (2007) model. Based on these p-values one would clearly reject the hypothesis that default times are generated by the estimated default intensities. After adding the macroeconomic variables suggested by Lando and Nielsen (2008), which is done in models III and IV, p-values go up. If the rating is not included (model III), the test no longer rejects for bin sizes 2 to 16. Thus, the results of Lando and Nielsen (2008) are broadly confirmed in our sample. With the rating included (model IV), the test again rejects for most bin sizes, but the p-values are larger than without the macroeconomic variables proposed by Lando and Nielsen (2008).

Adding the Baseline Component to the specification of the intensities tends to increase the p-values, in particular for smaller bin sizes. To see why it particularly effects smaller bin sizes, note that the intra-month timing does not affect a model's performance if a test is based on the defaults during an entire month. It only matters if a bin starts or ends with a fractional month. If the bin ends on a 16th, for example, the true expected default count would be almost equal to the default count over the entire month as most defaults occur before the 16th. If intra-month patterns are ignored, however, the expected default count for the time from the 1st to the 16th would be little more than one half of the expected monthly default count. The larger the bin size, the smaller is the weight of any fractional months at the beginning or the end of the bin compared to the full months within the bin, and the lower is their impact on the test.

With rolling estimation, p-values are moved away from rejection for models I, II and IV, which showed low p-values with fixed coefficients, and p-values are generally less dependent on the variables used in the intensity specification. Even with the choice of macro-variables suggested by Das et al. (2007) one would not reject for bin sizes one to six or one to eight (depending on whether the rating is included or not). The effect of adding the macro variables suggested by Lando and Nielsen (2008) is weaker than before; it only increases p-values for bin sizes 4, 10 and 16 (without Baseline Component). Moreover, adding the rating as is done in model IV no longer reduces the p-values. This could be due to a changing information content of ratings, which is captured through rolling estimation but which cannot be captured with fixed coefficients.

As before, adding the Baseline Component mostly leads to an increase of p-values. Overall, however, moving from an in-sample estimation to a rolling estimation has a stronger effect on p-values than the inclusion of the Baseline Component. Our findings are particularly important for practical applications, where one would typically use the most significant variables for the intensity specification. For example,

the rating has been found to be strongly significant in the model estimation of Section 3.1.3. We further found that the predictive power of a model with rating outperforms the one of a model without rating. Nevertheless, the rejection of a rating-based model (models II and IV) with fixed coefficients could mislead us to the conclusion that one would have incurred significant errors in predicting default clusters in the past implementing such a model. Our results with rolling estimation show that this is not the case.

Apart from Fisher's dispersion test, Das et al. (2007) also consider *Prahl's Test* (cf. Prahl (1999)) of clustered defaults. This test utilizes that aggregated intensities between the jumps of L are unit-exponentially distributed (cf. equation (3.3) if defaults are generated by the estimated intensities. Based on this test, Das et al. (2007) reject the H_0 hypothesis of correctly specified default intensities for the Duffie et al. (2007) model another time.

We do not consider this test because without further assumptions we doubt its significance for basically one reason: In our data set, there are many days at which several firms default. For instance, in the considered test period from 1985 on there are 490 defaults but only 416 different default dates. There are 60 dates at which more than one firm defaults and the maximum number of defaulted firms at a single day is 5. The question is: What is the aggregated intensity between default events at the same day? Not surprisingly, taking this intensity to be 0 by assuming that all firms jointly default at the end of this day would dramatically lower the chances of passing the test since we would obtain a large number of "0"-realizations of random variables, which are supposed to be unit-exponentially distributed. Another possibility would be to distribute the defaults equally across the date which would in connection with a re-estimated Baseline Component probably lead to sufficient results. Nonetheless, in either case, calculated p-values would depend to a possibly large extent on the way how we treat these defaults. Note further that the exponential distribution has been implicitly tested by testing for the Poisson distribution because this Poisson distribution is based on the same series of unit-exponentially distributed random variables which we would consider in Prahl's test.

Default Probabilities

After having explored the question of whether the path of observed *portfolio* defaults is likely to have been generated by the *aggregated* paths of default intensities, we now assess if estimated *single* default intensities provide default forecasts which are consistent with observed defaults on a *single* firm basis.

We base our analysis of model-implied default probabilities and defaults on a so-called *Spiegelhalter* test, see Tasche (2007). The test uses as input the default

probabilities of independent firms. In the setup considered here, conditional on a realization of the background process X^1 the firms are independent, and their conditional default probabilities are given as (cf. Proposition 2.5.1)

$$ {}_{T,t}q_i\left(x^1\right) := \mathbb{P}\left(t < \tau_i \leq T | \, \mathcal{F}_{T^*}^{X^1} \wedge \{\tau_i > t\}\right) = 1 - \exp\left(-\int_t^T \lambda_i(u, x^1(u))\right). $$

Furthermore, defaults of the portfolio objects in any two disjoint time intervals, $[t_1, T_1]$ and $[t_2, T_2]$ with $[t_1, T_1] \cap [t_2, T_2] = \emptyset$, are independent conditional on a realization of X^1. This can be utilized for a Spiegelhalter test to check whether firms default according to their estimated default probabilities

$$ {}_{T,t}q_i\left(r^1, r^2\right) = 1 - \exp\left(-\int_t^T \lambda(u, \hat{\theta}(u), r^1(u), r_i^2(u))du\right) $$

in the following way: Given our portfolio of I (conditionally) independent firms with (conditional) default probabilities ${}_{T,t}q_i\left(r^1, r_i^2\right)$ over the time period $[t, T]$, the Spiegelhalter test considers the *mean squared error* $\mathcal{M}$ between default probabilities and default indicators, i.e.

$$ \mathcal{M} = \frac{1}{I}\sum_{i=1}^{I}(1_{\tau_i \in [t,T]} - {}_{T,t}q_i\left(r^1, r_i^2\right))^2. $$

If defaults are independent and occur according to estimated default probabilities ${}_{T,t}q_i\left(r^1, r_i^2\right)$, we have that

$$ \begin{aligned} \mathbb{E}[\mathcal{M}] &= \frac{1}{I}\sum_{i=1}^{I} {}_{T,t}q_i\left(r^1, r_i^2\right)(1 - {}_{T,t}q_i\left(r^1, r_i^2\right)) \quad \text{and} \\ Var[\mathcal{M}] &= \frac{1}{I^2}\sum_{i=1}^{I} {}_{T,t}q_i\left(r^1, r_i^2\right)(1 - {}_{T,t}q_i\left(r^1, r_i^2\right))(1 - 2\,{}_{T,t}q_i\left(r^1, r_i^2\right))^2. \end{aligned} $$

Moreover, by the central limit theorem the random variable

$$ \frac{\mathcal{M} - \mathbb{E}\left[\mathcal{M}\right]}{\sqrt{Var\left[\mathcal{M}\right]}} $$

is asymptotically standard normally distributed. Based on this observation, we can conduct a test of the H_0 hypothesis *"single firms default according to the estimated default probabilities"* (and hence default intensities).

Since the test only uses default probabilities and not intensities, we have to specify a time grid $t_0 = 0 < t_1 < \ldots < t_K$ for which we want to consider the default probabilities of the single firms. In case of the equidistant time grid with grid size δ that we will use, the mean squared error which is actually used in the test is

$$ \mathcal{M} = \frac{1}{n^*}\sum_{k=1}^{K}\sum_{i=1}^{I}\left(1_{\delta(k-1)<\tau_i\leq\delta k} - {}_{\delta k,\delta(k-1)}q_i\left(r^1, r_i^2\right)\right)^2, $$

where n^* denotes the total number of observations. Thereby, a firm contributes to the MSE in period $(\delta(k-1), \delta k]$ if either all its variables are observed at the beginning of the period and the firm defaults during $(\delta(k-1), \delta k]$ or if all its variables are observed at the beginning and at the end / after the end of the period, i.e. if the firm has not defaulted during $(\delta(k-1), \delta k]$. It is worth mentioning that a firm usually enters several times but each time for a different time period; this does not impair the test since conditional on the path of X^1 defaults in disjoint time intervals are independent as previously pointed out.

In Table 3.4, the p-values of the Spiegelhalter test with respect to different grid sizes δ for fixed coefficients as well as for estimates derived with five-year rolling estimation windows are presented. The tests are performed for the defaults over the time period from January 1985 to April 2005 since for this period both estimates are available. In case of the considered one-year time grid, we assume that a firm's default intensity is constant during each time interval $(k-1, k]$ and given by the estimated intensity at the beginning of the period.

	Fixed Coefficients		Rolling Estimation	
	1-month grid size	1-year grid size	1-month grid size	1-year grid size
Model I	0.903	0.000***	0.040*	0.000***
Model II	0.915	0.000***	0.006**	0.000***
Model III	0.923	0.000***	0.019*	0.000***
Model IV	0.928	0.000***	0.004**	0.000***

Table 3.4: p-values of the Spiegelhalter test for equidistant time grids with sizes $\frac{1}{12}$ and 1. Intensity models I to IV differ in explanatory variables. Results are presented for fixed coefficients and rolling estimation. Low p-values would lead to a rejection of the hypothesis of well-specified intensities on a single firm basis.

The presented results for the one-year grid size show that current default intensities do not provide a good proxy for the single default probabilities over the subsequent year – no matter which model specification or estimation approach we consider. This has to be attributed to the decreasing predictive power of the intensities. As previously mentioned, a default intensity only provides a firm's default probability over the next infinitesimally small time step.

Regarding the one-month time grid, we find that the Spiegelhalter test is generally more favorable for the models with fixed coefficients: The p-values of all four specifications are about 90% in this case. With rolling estimations, p-values are

considerably lower leading to a rejection of all estimated models at the 5% significance level. One possible explanation of this outcome is that the total estimated aggregated intensity is larger than the number of observed defaults, while with fixed coefficients the aggregated intensity matches the number of observed defaults.[8] If the observed number of defaults is smaller than the expected number, the observed mean squared error $\mathcal{M}$ tends to be smaller than its estimated expectation; the smaller the observed mean squared error $\mathcal{M}$ is in comparison with its expectation, the more likely becomes a rejection of the estimated model on a single firm basis. The results of the Spiegelhalter test therefore show that in case of rolling estimations at least some of the default intensities might not be consistent with the observed default behavior on a single firm basis. However, the fact that switching between different explanatory variables (models I-IV) has only a minor effect on the results, indicates that – without further investigations – these findings have to be treated precautiously. This provides a new area of future research.

In summary, we conclude that Baseline Component and rolling estimations considerably improve the estimation results for the default intensities on a portfolio basis. When both modifications are made, the test suggested by Das et al. (2007) no longer rejects the hypothesis of a well-specified aggregated portfolio intensity.

In a recent contribution, Lando and Nielsen (2008) arrive at similar results regarding the aggregated portfolio intensity. Specifically, they modify the intensity model of Duffie et al. (2007) by adding explanatory variables. In our data set, adding industrial production and the term spread as suggested by Lando and Nielsen (2008) also renders the model forecasts consistent with observed default clustering. As long as estimation is conducted in-sample, however, the results are sensitive to the inclusion of the credit rating, which is a natural candidate for inclusion because it significantly improves standard metrics of default prediction power. Taken together, the results presented by Lando and Nielsen (2008) and ourselves provide a strong case for reconsidering the results of Das et al. (2007).

3.2 A Model for Default Intensities

In the previous Section 3.1, we estimated default intensities for a large number of US and non-US corporates and assessed the quality of our estimations with respect to default prediction and default clustering. Further investigations regarding the dynamics of the estimated intensities will follow in Sections 3.3 and 3.5. By introducing a model for the joint intensity dynamics, the current section builds the foundation of this analysis.

[8]For example, in case of model IV with rolling estimations the aggregated integrated intensity is 568.70 compared with only 490 observed defaults.

The model that we introduce in the following is inspired by Duffie and Gârleanu (2001) and Mortensen (2006) but also includes our findings presented in Section 2.8, where we gave a detailed discussion of the Duffie and Gârleanu (2001) and its potential shortcomings. By introducing additional features, which we consider to be important, we obtain a model that represents an extension of their model. As already mentioned, the performance of the different model versions in the context of actual and simulated default (intensity) data is then assessed in Sections 3.3, 3.4 and 3.5.

Introduction of Our Intensity Model

Like the Duffie and Gârleanu (2001) model, our model also represents a Bottom-Up model and relies on the conditional independence assumption (see Subsection 2.5.1). Also, we use the same idea as in Mortensen (2006) of one common factor λ^c which all intensities are exposed to and propose default-trigger processes Λ_i of the form[9]

$$\Lambda_i(t) = a_i \left(\int_0^t \lambda^c(s)ds + \int_0^t \tilde{\lambda}_i(s)ds \right).$$

In terms of Definition 2.1.1, we assume the following:

- The background process X^1 is given by

  $$X^1 = \left(\lambda^c, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_I, Z \right),$$

  where the regime process Z denotes a Markov-Chain, i.e. a finite state Markov process with constant transition intensities. We denote by $\overline{Z}$ its initial state and by $\mathcal{Q}$ its generator matrix containing the transition intensities. Moreover, the r-dimensional state space of Z is as in Section 2.7 given by $\{e_1, \ldots, e_r\}$ with $e_m = (1_{1=m}, 1_{2=m}, \ldots, 1_{r=m})$ for all $m \in \{1, \ldots, r\}$. For simplicity, we assume that Z evolves independently of $\left(\lambda^c, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_I \right)$, although extensions with dependencies between both processes are straight-forward. The remaining, first $I+1$ coordinates of the background process X^1 are assumed to be positive processes, which are not necessarily independent (for details see below). As always, we assume that $\mathcal{F}_{T^*}^{X^1} = \mathcal{F}_{T^*}^{\Lambda}$ and $\mathcal{F}_t^N$ are independent given $\mathcal{F}_t^{X^1}$.

- $b_i(s, X^1(s), X^2(s)) = a_i \left(\lambda^c(s) + \tilde{\lambda}_i(s) \right)$.

- $\nu_m(s, \mathbb{R}_{0+}^I) = 0$, for each m.

It is worth noting that simultaneous defaults are not possible in our model since $\nu_m(s, \mathbb{R}_{0+}^I) = 0$ (cf. Proposition 2.3.1).

[9] As usual, generalizations to p-dimensional factor processes are straight-forward.

Apart from the regime process Z, so far our model is the same as in Mortensen (2006). Both models, however, differ in the dynamics which are presumed for $\left(\lambda^c, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_I\right)$. We assume that each λ^c and $\tilde{\lambda}_i$ evolves according to the SDE

$$d\lambda(t) = \kappa\left(Z(t)\varpi\eta - \lambda(t)\right)\, dt + \sigma\sqrt{\lambda(t)}\, dW(t) + dJ(t), \quad \lambda(0) = \overline{\lambda}. \tag{3.4}$$

Here $\varpi \in [0,1]^r$ and κ, σ and η are positive constants, W is a standard Wiener process and J is now – different to the BAJD model considered in Section 2.8 – a point process that jumps with intensity

$$Z(t)\varpi\mu + \xi^{(1)}\lambda(t) + \xi^{(2)}\Theta(t) \tag{3.5}$$

and has positive, independently $\mathrm{Exp}\left(\frac{1}{\zeta}\right)$-distributed jumps. The process Θ that shows up in the jump intensity specification is given by the SDE

$$d\Theta(t) = \left(\lambda(t) - \epsilon\Theta(t)\right) dt, \quad \Theta(0) = \overline{\Theta} \tag{3.6}$$

with solution

$$\Theta(t) = \overline{\Theta} + \int_0^t \lambda(s) e^{-\epsilon(t-s)} ds. \tag{3.7}$$

Furthermore, $\mu, \zeta, \xi^{(1)}, \epsilon$ and $\xi^{(2)}$ are positive constants. We will call a process λ of the form given by equations (3.4), (3.5) and (3.6) a *basic jump diffusion* and denote it by

$$BJD\left(\overline{\lambda}, \overline{\Theta}, \overline{Z}, \mathcal{Q}, \varpi, \eta, \kappa, \sigma, \mu, \zeta, \xi^{(1)}, \epsilon, \xi^{(2)}\right).$$

Having defined the marginal dynamics of each λ^c and $\tilde{\lambda}^i$ in our model, it still remains to specify their joint dynamics. To do so, we presume that

$$\begin{array}{lll} \lambda^c & \text{is a} & BJD\left(\overline{\lambda}, \overline{\Theta}, \overline{Z}, \mathcal{Q}, \varpi, \eta, \kappa, \sigma, \mu, \zeta, \xi^{(1)}, \epsilon, \xi^{(2)}\right) \quad \text{and} \\ \tilde{\lambda}_i & \text{is a} & BJD\left(\overline{\lambda}, \overline{\Theta}, \overline{Z}, \mathcal{Q}, 1-\varpi, \eta, \kappa, \sigma, \mu, \zeta, \xi^{(1)}, \epsilon, \xi^{(2)}\right). \end{array} \tag{3.8}$$

In addition, we assume that λ^c and $\tilde{\lambda}_i$ are driven by independent Wiener processes W and jump processes J that are assumed to be orthogonal. This means that conditional on the path of the regime process Z the λ^c and $\tilde{\lambda}_i$ are independent.

Like in the Duffie and Gârleanu (2001) model investigated in Section 2.8, ϖ controls the dependence level between the portfolio objects. Values close to 1 imply a strong dependence between the intensities, while values close to 0 entail that intensities evolve almost independently. However, the dependence level now depends on the regime of Z, too, because the state of Z identifies one of the coordinates of ϖ as the current dependence level. Our model can therefore incorporate times with low dependencies and times with strong dependencies between the firms' intensities.

The probability for a change in the dependence level depends on the generator matrix $\mathcal{Q}$. It is worth noting that the Duffie and Gârleanu (2001) model is included as a special case in our model; it is obtained by assuming only one regime for Z, i.e. $r = 1$, and by setting $\xi^{(1)} = \xi^{(2)} = 0$.

Model Properties and Important Formulas

Before analyzing the marginal intensity dynamics in our model, we first state some important "calculation rules", which apply when working with basic jump diffusions. Also, these rules will be very useful in Section 3.5. Given two basic jump diffusions λ^1 and λ^2 with parameters $BJD\left(\overline{\lambda}_k, \overline{\Theta}_k, \overline{Z}, \mathcal{Q}, \varpi_k, \eta, \kappa, \sigma, \mu_k, \zeta, \xi^{(1)}, \epsilon, \xi^{(2)}\right)$ and $k \in \{1, 2\}$, which are independent conditional on the path of the regime process Z, we have that

$$\lambda^1 + \lambda^2 \quad \text{is a} \quad BJD\left(\overline{\lambda}_1 + \overline{\lambda}_2, \overline{\Theta}_1 + \overline{\Theta}_2, \overline{Z}, \mathcal{Q}, \varpi_1 + \varpi_2, \eta, \kappa, \sigma, \mu_1 + \mu_2, \zeta, \xi^{(1)}, \epsilon, \xi^{(2)}\right) \tag{3.9}$$

$$a\lambda_k \quad \text{is a} \quad BJD\left(\overline{\lambda}_k, \overline{\Theta}_k, \overline{Z}, \mathcal{Q}, \varpi_k, a\eta, \kappa, \sqrt{a}\sigma, \mu_k, a\zeta, \frac{\xi^{(1)}}{a}, \epsilon, \frac{\xi^{(2)}}{a}\right). \tag{3.10}$$

Both results can easily be shown using the Itô-formula. We use these rules to investigate the marginal dynamics of the default intensities in our model. Namely, an application of equations (3.9) and (3.10) yields that a firm's default intensity λ_i, i.e.

$$\lambda_i := a_i(\lambda^c + \tilde{\lambda}_i),$$

evolves according to the SDE

$$\begin{aligned} d\lambda_i(t) &= \kappa\left(a_i\eta - \lambda_i(t)\right)\, dt + \sqrt{a_i}\sigma\sqrt{\lambda_i(t)}\, dW(t) + dJ(t), \\ \lambda_i(0) &= \overline{\lambda}_i = a_i\left(\overline{\lambda}^c + \overline{\tilde{\lambda}}_i\right), \end{aligned} \tag{3.11}$$

where J jumps with intensity

$$\mu + \frac{\xi^{(1)}}{a_i}\lambda(t) + \frac{\xi^{(2)}}{a_i}\Theta_i(t) \tag{3.12}$$

and has positive, independently $\mathrm{Exp}\left(\frac{1}{a_i\zeta}\right)$-distributed jumps. Θ_i is again given by the SDE

$$d\Theta_i(t) = \left(\lambda_i(t) - \epsilon\Theta_i(t)\right) dt, \quad \Theta_i(0) = a_i\left(\overline{\Theta}^c + \overline{\tilde{\Theta}}_i\right). \tag{3.13}$$

Observe that the dynamics of the single λ_i do not depend on the regime process Z anymore. Since Z only affects their joint dynamics, it can therefore be interpreted as the process that drives the portfolio dependence structure. In the following, we

call a process λ_i of the form given by equations (3.11), (3.12) and (3.13) a *self-affecting affine jump diffusion with memory effect* (for an explanation of this name see below) and denote it by

$$SAJDM\left(\overline{\lambda}_i, \overline{\Theta}_i, a_i\eta, \kappa, \sqrt{a_i}\sigma, \mu, a_i\zeta, \frac{\xi^{(1)}}{a_i}, \epsilon, \frac{\xi^{(2)}}{a_i}\right).$$

As previously indicated, our credit portfolio model includes other models as special cases. In particular, the single intensity version of the model, given by equations (3.11), (3.12) and (3.13), includes:

- $\mu = \xi^{(1)} = \xi^{(2)} = 0$: *Cox-Ingersoll-Ross* model (CIR). Model without jumps. This process has been introduced by Cox et al. (1985) as a model of the short rate.
- $\xi^{(1)} = \xi^{(2)} = 0$: *Basic affine jump diffusion* (BAJD). Jumps occur independently of the state of λ_i. This model has been proposed by Duffie and Gârleanu (2001) as a model of the default intensity within a credit portfolio application (see Section 2.8).
- $\xi^{(2)} = 0$: The jump intensity of J is state dependent but does not depend on the process Θ_i. We refer to such a model as a *self-affecting affine jump diffusion* (SAJD) in the following.

As the main difference between our model and the well-known CIR and BAJD models, the jump intensity of J now depends on the state of λ_i as well as on a process Θ_i. This process keeps hold of the past intensity, where more weight is put on the more recent history (cf. equation (3.7)). How strong past intensities enter depends on the parameter ϵ. For ϵ very large, for example, only the very recent intensity history will be relevant. In summary, jumps of the intensity increase the probability of further jumps in our model, and high intensities in the past imply a high likelihood for future intensity jumps. This means that periods of high intensity volatility are likely to be followed by a high activity of the intensity in the future resulting in "volatility explosions".

After having analyzed the marginal intensity dynamics in our model, we now return to the portfolio perspective. First, we apply Theorem 2.7.1 to calculate an important transform of an integrated basic jump diffusion:

Proposition 3.2.1 *If the process λ evolves according to equations (3.4), (3.5) and (3.6), we have that*

$$\begin{aligned} {}_{T,t}\varphi_{\int_0^\bullet \lambda(s)ds}(c) &= E\left[e^{-c\int_t^T \lambda(s)ds}\middle|\, \lambda(t), \Theta(t), Z(t)\right] \\ &= e^{Z(t)u(t,T)+v_1(t,T)\lambda(t)+v_2(t,T)\Theta(t)}, \end{aligned} \tag{3.14}$$

where c is of the form

$$c = \begin{cases} -iw & , \quad w \in \mathbb{R} \quad or \\ w & , \quad w \in \mathbb{R}_{0+}. \end{cases}$$

Furthermore, the complex-valued functions u, v_1 and v_2 of equation (3.14) solve the following system of ODEs: For all $m \in \{1, \ldots, r\}$, u_m solves

$$0 = \dot{u}_m + v_1 \kappa \eta \varpi_m + \mu \varpi_m \frac{\zeta v_1}{1 - \zeta v_1} + \sum_{n \neq m} \left(e^{u_n - u_m} - 1 \right) \mathcal{Q}_{mn},$$

and v_1 and v_2 solve

$$\begin{aligned} 0 &= \dot{v}_1 - c - v_1 \kappa + v_2 + 0.5 v_1^2 \sigma^2 + \xi^{(1)} \frac{\zeta v_1}{1 - \zeta v_1}, \\ 0 &= \dot{v}_2 + \xi^{(2)} \frac{\zeta v_1}{1 - \zeta v_1} - \epsilon v_2 \end{aligned}$$

with terminal conditions $u(T,T) = v_1(T,T) = v_2(T,T) = 0$.

Proof: The result is a special case of Theorem 2.7.1 and the ODEs stated there. Here, we apply the theorem to $Y' = (\int_0^\bullet \lambda(s) ds, \lambda, \Theta)$ with $(c_1', c_2', c_3') = (c, 0, 0)$. First, we get that $v_1'(t,T) \equiv -c$ and can therefore plug $v_1'(t,T) \equiv -c$ into the ODEs for v_2' and v_3'. By finally setting $v_1(t,T) = v_2'(t,T)$ and $v_2(t,T) = v_3'(t,T)$, the claim follows.

□

Since the SAJDM model represents a special case of the BJD model, Proposition 3.2.1 can be used to calculate single survival probabilities. These are obtained by calculating the transform (3.14) for $c = 1$ and by setting the term which is related to jumps of the regime process Z to 0, because in this case the intensity dynamics do not depend on Z. Then, u, v_1 and v_2 are given as solutions of

$$\begin{aligned} 0 &= \dot{u} + v_1 \kappa \eta + \mu \frac{\zeta v_1}{1 - \zeta v_1}, \\ 0 &= \dot{v}_1 - 1 - v_1 \kappa + v_2 + 0.5 v_1^2 \sigma^2 + \xi^{(1)} \frac{\zeta v_1}{1 - \zeta v_1}, \\ 0 &= \dot{v}_2 + \xi^{(2)} \frac{\zeta v_1}{1 - \zeta v_1} - \epsilon v_2, \end{aligned}$$

with terminal conditions $u(T,T) = v_1(T,T) = v_2(T,T) = 0$. As we will show in Section 3.5, Proposition 3.2.1 will be also very useful for calculating the portfolio loss distribution.

By analyzing the expected volatility (cf. Definition 2.6.1 on p. 50) of the Duffie and Gârleanu (2001) model in Section 2.8, we found that defaults are relatively evenly distributed over time in this model. In our model, the expected volatility can similarly be calculated as

$$
\begin{aligned}
{}_{T,t}\sigma_L \;=\; & \mathbb{E}\left[\sum_{i=1}^{I}\int_t^T a_i 1_{\tau_i>s}\left(\lambda^c(s)+\tilde{\lambda}_i(s)\right)ds\,\middle|\, X^1(t)\right] \\
& +\mathbb{E}\left[\int_t^T\left(\left(\sum_{i=1}^{I}1_{\tau_i>s}a_i\right)^2\sigma^2\lambda^c(s)+\sum_{i=1}^{I}1_{\tau_i>s}a_i^2\sigma^2\tilde{\lambda}_i(s)\right)ds\right. \\
& +2\zeta^2\int_t^T\left(Z(s)\varpi\mu+\xi^{(1)}\lambda^c(s)+\xi^{(2)}\Theta^c(s)\right)\left(\sum_{i=1}^{I}1_{\tau_i>s}a_i\right)^2 ds \\
& \left.+2\zeta^2\sum_{i=1}^{I}\int_t^T\left(Z(s)(1-\varpi)\mu+\xi^{(1)}\tilde{\lambda}_i(s)+\xi^{(2)}\tilde{\Theta}_i(s)\right)1_{\tau_i>s}a_i^2 ds\,\middle|\, X^1(t)\right] \\
& +\mathbb{E}\left[\int_t^T\sum_{i=1}^{I}1_{\tau_i>s}a_i^3\left(\lambda^c(s)+\tilde{\lambda}_i(s)\right)^3 ds\,\middle|\, X^1(t)\right].
\end{aligned}
$$

The first term, the second and the last term are equal in the Duffie and Gârleanu (2001) model. What is different are the two terms in the middle, which are related to the jump behavior of the intensities. Now, both terms depend on the state of the intensity as well as on the averaged, weighted past intensity. This shows that in our model the default clustering will vary much more over time than in the Duffie and Gârleanu (2001) model.

Figure 3.7 displays simulated paths of a BAJD and a SAJD model. While in the first model intensity jumps are evenly distributed over time, in the SAJD model they are clustered (in the example path around year 3). Let us, for instance, assume that the displayed paths represent the factor λ^c in our portfolio model. Then, in the SAJD model defaults would be heavily clustered around year 3, while in the BAJD model the observed default clustering would be less pronounced.

All in all, the introduced model represents a flexible approach for modeling the dynamics of default intensities. From a mathematical point of view, the analytical tractability of the model is a further appealing and important feature. A detailed discussion of our model choice in the context of actual default intensity data is postponed to Sections 3.3 and 3.5 where we present its calibration to the default intensities estimated in Section 3.1.

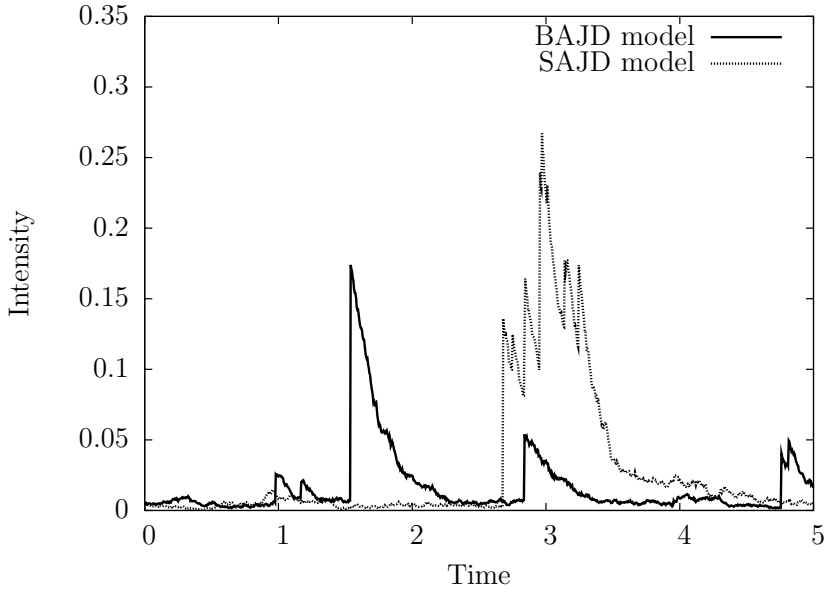


Figure 3.7: Simulated paths of a BAJD and a SAJD model over a five-year period. The parametrization of the diffusion component is in both models $\eta = 0.005, \kappa = 5.5$ and $\sigma = 0.1$. For the BAJD model, the jump parameters are $\mu = 0.975$ and $\zeta = 0.07$, while for the SAJD model we have $\mu = 0, \zeta = 0.07$ and $\xi^{(1)} = 60$. Both models imply a five-year default rate of about 8%.

3.3 Single Firm Modeling

We now consider the estimation of the different versions of our model introduced in the previous section to the intensities estimated in Section 3.1. While Section 3.5 addresses the challenge of modeling the *joint* dynamics of the estimated default intensities, in this section we focus on modeling default intensities on a single firm basis. In particular, we deal with the following questions:

- *What models describe the single intensity dynamics best from a statistical point of view?*
- *How do the different models perform with respect to default prediction?*

Our analysis uses estimated intensities based on model IV with Baseline component and rolling estimation windows of five years and also – as a robustness check – intensities based on model III with fixed coefficients and Baseline component since both models where not rejected by Fisher's dispersion test conducted in Section 3.1 (cf. Table 3.3 on p. 96).

The remainder of the section is organized as follows. Subsection 3.3.1 describes our

methodology. In Subsection 3.3.2, we then present the estimation results, compare the statistical significance of the different models and assess their predictive power.

3.3.1 Model Properties and Estimation Methodology

Table 3.5 summarizes model-implied instantaneous means and variances of default intensities for the four single intensity models (CIR, BAJD, SAJD, SAJDM), which we will consider throughout this section. It shows that a default intensity following a SAJD or SAJDM model allows for stronger "volatility explosions" than the traditional models. Although in case of a CIR or BAJD model the volatility $\sigma\sqrt{\lambda(t)}$ is also state dependent, very fast, sudden increases of the intensities are less likely. Making the jump intensity of J state dependent and dependent on the averaged, weighted past intensity represents a feasible way of enriching the established models by such effects.

i) Mean

	CIR	BAJD	SAJD
$\frac{E_t[d\lambda(t)]}{dt}$	$\kappa(\eta-\lambda(t))$	$\kappa(\eta-\lambda(t))+\mu\zeta$	$\kappa(\eta-\lambda(t))+\zeta(\mu+\xi^{(1)}\lambda(t))$

	SAJDM
$\frac{E_t[d\lambda(t)]}{dt}$	$\kappa(\eta-\lambda(t))+\zeta(\mu+\xi^{(1)}\lambda(t)+\xi^{(2)}\Theta(t))$
$\frac{E_t[d\Theta(t)]}{dt}$	$\lambda(t)-\epsilon\Theta(t)$

ii) Variance

	CIR	BAJD	SAJD
$\frac{Var_t[d\lambda(t)]}{dt}$	$\sigma^2\lambda(t)$	$\sigma^2\lambda(t)+\mu 2\zeta^2$	$\sigma^2\lambda(t)+2\zeta^2(\mu+\xi^{(1)}\lambda(t))$

	SAJDM
$\frac{Var_t[d\lambda(t)]}{dt}$	$\sigma^2\lambda(t)+2\zeta^2(\mu+\xi^{(1)}\lambda(t)+\xi^{(2)}\Theta(t))$
$\frac{Var_t[d\Theta(t)]}{dt}$	0

Table 3.5: Instantaneous mean and variance of the default intensity λ in a CIR, BAJD, SAJD or SAJDM single firm intensity model.

Our data set comprises 331954 (monthly) default intensities for 3846 different firms which have been estimated in Section 3.1. For the time being, we are only interested in calibrating the four different models to the default intensity time series of

each single firm. Unfortunately, our data set contains firms for which only few observations are available. To reduce variances of the estimators on the one hand but to diminish the data set not too strongly on the other, we disregard firms with less than 24 observations. This constraint reduces the data set to 250090 observations belonging to 3241 different firms.

Model inference is conducted via ML. Given that we observe a firm's default intensity λ on a time grid $t_1 < \ldots < t_K$, – due to the Markovian model structure – the corresponding likelihood function $\mathcal{L}(\theta;\lambda)$ is given by

$$\mathcal{L}(\theta;\lambda) = \prod_{k=2}^{K} {}_{t_k,t_{k-1}}g_\lambda\left(\lambda(t_k);\theta\right), \tag{3.15}$$

where ${}_{t+\Delta,t}g_\lambda$ denotes the conditional density of $\lambda(t+\Delta)$ given $\lambda(t)$ and a particular model parametrization θ. As usual, the ML estimator $\hat{\theta}_i$ is obtained as

$$\hat{\theta}_i = \arg\max_{\theta} \ \mathcal{L}(\theta;\lambda).$$

In the most general case of a SAJDM model, the parameter vector θ has 8 entries, i.e. $\theta = \left(\eta,\kappa,\sigma,\mu,\eta,\xi^{(1)},\xi^{(2)},\epsilon\right)^T$. The main difficulty in the ML estimation is to derive the transition density ${}_{t+\Delta,t}g_\lambda$ for the intensity models considered.

Since all models represent exponential affine models, a possible way of obtaining this density is the Fourier inversion of the corresponding conditional characteristic function of $\lambda(t+\Delta)$ given $\lambda(t)$:

$$\begin{aligned} {}_{t+\Delta,t}\varphi_\lambda(-iw) &= \mathbb{E}\left[e^{iw(\lambda(t+\Delta)-\lambda(t))}\middle|\,\lambda(t),\Theta(t)\right] = \\ &= e^{(u(t,t+\Delta)+(v_1(t,t+\Delta)-iw)\lambda(t)+v_2(t,t+\Delta)\Theta(t))}. \end{aligned} \tag{3.16}$$

It can be calculated based on Theorem 2.7.1 and is given semi-analytically in our model up to the solution of an ODE system. Here, the complex-valued functions u, v_1 and v_2 solve the following system of ODEs

$$\begin{aligned} 0 &= \dot{u} + v_1\kappa\eta + \mu\frac{\zeta v_1}{1-\zeta v_1}, \\ 0 &= \dot{v}_1 - v_1\kappa + v_2 + 0.5v_1^2\sigma^2 + \xi^{(1)}\frac{\zeta v_1}{1-\zeta v_1}, \\ 0 &= \dot{v}_2 + \xi^{(2)}\frac{\zeta v_1}{1-\zeta v_1} - \epsilon v_2, \end{aligned} \tag{3.17}$$

with terminal conditions $u(t+\Delta,t+\Delta) = 0 = v_2(t+\Delta,t+\Delta)$ and $v_1(t+\Delta,t+\Delta) = iw$. In case of a BAJD model ($\xi^{(1)} = \xi^{(2)} = 0$), even explicit solutions of the ODEs can be stated, for details see Appendix A. As hinted by Singleton (2001), ${}_{t+\Delta,t}g_\lambda(x)$ can then be calculated by Fourier inversion:

$${}_{t+\Delta,t}g_\lambda(x;\theta) = \frac{1}{\pi}\int_{\mathbb{R}_+} Re\left[e^{-iwx}\ {}_{t+\Delta,t}\varphi_\lambda(-iw)\right]dw,$$

where *Re* denotes the real part of complex numbers. For a detailed discussion of ML estimation based on the characteristic function we refer to Singleton (2001).

However, in view of our large data set with 250090 observations inverting the characteristic function is computationally too cumbersome since for each point of each intensity time series, the characteristic function has to be inverted. Therefore, instead of applying ML estimation based on the characteristic function we approximate the conditional density ${}_{t+\Delta,t}g_\lambda(\cdot;\theta)$ and estimate parameters based on this approximation. ML estimates derived from an inversion of the characteristic function are only used to check the quality of the approximation.

The basis of our approximation of ${}_{t+\Delta,t}g_\lambda(\cdot;\theta)$ is the time-discretization of the SDE (3.11), which yields

$$\lambda(t+\Delta)-\lambda(t)=\kappa(\eta-\lambda(t))\Delta+\sigma\sqrt{\lambda(t)\Delta}N_{0,1}+\Upsilon\Delta J(t) \tag{3.18}$$

with $N_{0,1}$ denoting a standard-normally random variable. Υ is given as an $Exp\left(\frac{1}{\zeta}\right)$-distributed random variable and $\Delta J(t)$ is Bernoulli-distributed with parameter $1-exp(-\Delta(\mu+\xi^{(1)}\lambda(t)+\xi^{(2)}\Theta(t)))$. Based on this discretization, the approximate conditional density ${}_{t+\Delta,t}\tilde{g}_\lambda(\cdot;\theta)$ is given by a convolution of a Gaussian and a mixture of an exponential and a Bernoulli-distributed random variable. In particular, given $\Delta J(t)=1$ and given $\lambda(t)$, the likelihood of observing $\lambda(t+\Delta)$ is

$$h(o^{(1)},o^{(2)}):=\int_0^\infty \frac{1}{\zeta}e^{-\frac{x}{\zeta}}\frac{1}{\sqrt{2\pi}o^{(2)}}e^{-0.5\left(\frac{o^{(1)}-x}{o^{(2)}}\right)^2}dx, \tag{3.19}$$

where $o^{(1)}:=\lambda(t+\Delta)-\lambda(t)-\kappa(\eta-\lambda(t))\Delta$ and $o^{(2)}:=\sigma\sqrt{\lambda(t)\Delta}$. Therefore the approximate transition likelihood is given by

$$\begin{aligned}{}_{t+\Delta,t}\tilde{g}_\lambda(\lambda(t+\Delta);\theta)=&\left(1-e^{-\Delta\left(\mu+\xi^{(1)}\lambda(t)+\xi^{(2)}\Theta(t)\right)}\right)h(o^{(1)},o^{(2)}) \\ &+e^{-\Delta\left(\mu+\xi^{(1)}\lambda(t)+\xi^{(2)}\Theta(t)\right)}f_{N_{\kappa(\eta-\lambda(t))\Delta,\sigma\sqrt{\lambda(t)\Delta}}}(\lambda(t+\Delta)-\lambda(t)),\end{aligned} \tag{3.20}$$

where $f_{\epsilon_1,\epsilon_2}$ denotes the density of a Gaussian random variable with mean ϵ_1 and standard deviation ϵ_2. It is important to note that the integral (3.19) can substantially be simplified by considering integration with respect to

$$z:=\frac{1}{o^{(2)}}\left(x-\left(o^{(1)}-\frac{o^{(2)}o^{(2)}}{\zeta}\right)\right)$$

instead of x, which yields

$$\begin{aligned}
&\int_0^\infty \frac{1}{\zeta} e^{-\frac{x}{\zeta}} \frac{1}{\sqrt{2\pi} o^{(2)}} e^{-0.5\left(\frac{o^{(1)}-x}{o^{(2)}}\right)^2} dx \\
&\quad = \frac{1}{\zeta} e^{-\frac{o^{(1)}}{\zeta}+\frac{o^{(2)}o^{(2)}}{2\zeta^2}} \int_{-\frac{1}{o^{(2)}}\left(o^{(1)}-\frac{o^{(2)}o^{(2)}}{\zeta}\right)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-0.5z^2} dz \\
&\quad = \frac{1}{\zeta} e^{-\frac{o^{(1)}}{\zeta}+\frac{o^{(2)}o^{(2)}}{2\zeta^2}} \left(1 - F_{N_{0,1}}\left(-\frac{o^{(1)}}{o^{(2)}} + \frac{o^{(2)}}{\zeta}\right)\right).
\end{aligned}$$

Here, $F_{N_{0,1}}$ denotes the CDF of a standard Gaussian random variable. We already pointed out that computational efficiency is crucial to our problem due to the large data set considered. After our integral transformations – instead of calculating the integral (3.19) by brute force numerical integration – the likelihood can be very efficiently computed by evaluating a standard Gaussian CDF and by multiplying it with some factor. Still, there remains one problem: For small values of ζ, e.g. $\zeta < 0.01$, the factor becomes very large while the CDF approaches 0. To avoid numerical instabilities, we rely in this case on Gauss-Laguerre integration (see e.g. Press et al. (2007)) of the basic integral (3.19). We found that using 10 points in the Gauss-Laguerre integration is sufficient to guarantee a high degree of accuracy, in particular a higher degree of accuracy than provided by the numerical integration routines implemented in the *GNU Scientific Library* (GSL).[10] On the other hand, integration still remains faster by approximately a factor of 100 compared to brute force numerical integration based on the GSL. In total, the approximation of the SDE in combination with the efficient calculation of the integral (3.19) gives rise to a very fast ML estimation algorithm.

3.3.2 Estimation Results for the Single Firm Intensity Models

In the following, we provide the results of our estimation. First we assess the goodness of the approximation (3.18) of the transition density in the context of two extreme example default intensity paths, which can be found in the data set. We then investigate the ability of the different models to explain the default intensities of all firms. Finally, we measure the predictive power of the intensity models.

Goodness of the Approximation

In Figure 3.8 the estimated default intensities of two firms in the data set, General Electric and Waxman Industries, are displayed. Both firms represent extreme de-

[10] See www.gnu.org for detailed information.

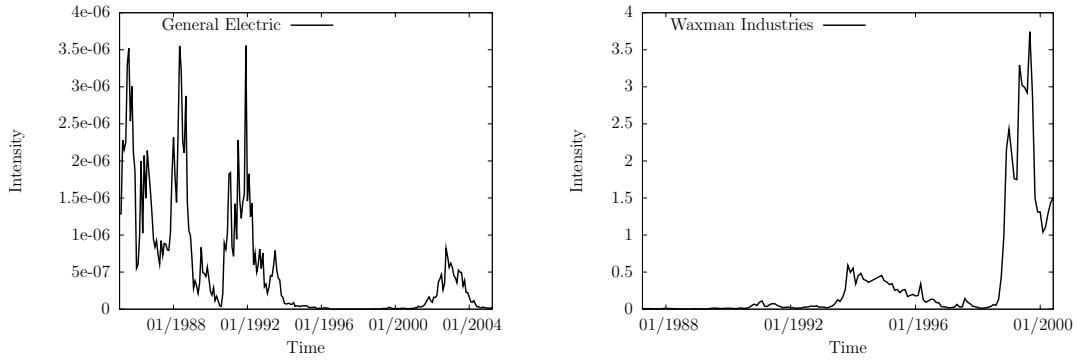


Figure 3.8: Estimated default intensities of General Electric (left) and Waxman Industries (right) based on model IV with Baseline Component and rolling five-year estimation windows.

fault intensity examples: While General Electric exhibits one of the lowest default intensities in the data set with values of magnitude 10^{-5}, Waxman Industries's maximum observed intensity is close to 4 and the firm defaulted at the end of the time series in June 2000 after a missed coupon payment. It is worth mentioning that a default intensity of 10^{-5} implies a monthly default probability of 0.0000833% whereas an intensity of 4 corresponds to a monthly default probability of more than 28%.

	η	κ	σ	μ	ζ	$\xi^{(1)}$
General Electric						
CIR Model	2.92e-08	0.118	0.00115	–	–	–
	(1.71e-07)	(0.314)	(0.00136)	–	–	–
Waxman Industries						
CIR Model	100	0.000278	0.768842	–	–	–
	(1.03e-07)	(1.85e-06)	(0.338)	–	–	–
BAJD Model	0.0247	1.16	0.383	1.00	0.499	–
	(1.43e-12)	(0.911)	(0.291)	(1.32)	(0.401)	–
SAJD Model	0.0173	2.38	0.316	0.0	0.345	17.9
	(0.0320)	(2.83)	(0.290)	(0.0)	(0.422)	(22.7)

Table 3.6: Parameter estimates of three models (CIR, BAJD, SAJD) for General Electric and Waxman Industries based on the approximation of the transition density. For General Electric the jump part has been estimated to be 0; for Waxman Industries, the memory effect was insignificant. Estimates based on the characteristic function are provided in parentheses.

Parameter estimates for the two example firms are presented in Table 3.6. We do not report estimates of the jump models in case of General Electric since the jump component was estimated to be 0. Furthermore, in case of Waxman Industries the memory effect was insignificant. Estimates obtained from an inversion of the characteristic function are provided in parentheses. By simply looking at Figure 3.8, one would have probably guessed that a jump-diffusion model is by far better able to explain the observed intensity path of Waxman Industries. This conjecture is underpinned by our estimates. For the CIR, estimates are rather "dubious", but as soon as we include the possibility of jumps estimated parameters become much more plausible: The mean reversion level drops, the speed of mean reversion increases and volatility moves from the Brownian part into the jump part. In particular, when incorporating a self-affecting model feature jumps are estimated to be of pure self-affecting nature since μ is 0.

Furthermore, when comparing the estimates based on our approximation with the results derived from an inversion of the characteristic function we observe that the values are similar aside from the CIR estimates in case of Waxman Industries. However, in this case, the CIR model is obviously not capable of modeling the intensity path sufficiently well probably leading to unstable estimators. All in all, the proposed approximation seems to give rise to reliable estimates.

Comparing the Statistical Significance of the Different Models

The examples of Waxman Industries and General Electric show that there are firms in the data set whose default intensities are strongly driven by jumps (Waxman Industries) and other firms whose default intensity dynamics seem to be sufficiently well described by a simple CIR process (General Electric). In this subsection, we formalize our investigation of the question of which of the four models appears to be most appropriate in order to model the default intensity dynamics.

The typical way of comparing nested models estimated by ML is to apply a χ^2-significance test. By standard ML theory,

$$-2\left(\log\left(\mathcal{L}_0(\hat{\theta}_0)\right) - \log\left(\mathcal{L}_1(\hat{\theta}_1)\right)\right)$$

is asymptotically χ^2-distributed with $n_1 - n_0$ degrees of freedom. Here, $\mathcal{L}_0$ ($\mathcal{L}_1$) denotes the likelihood function under H_0 (H_1) and n_0 (n_1) the number of parameters under H_0 (H_1).

An alternative approach, which is sometimes considered in the literature and does not depend on large sample theory, is to consider the so-called *Bayes factor*, see Kass and Raftery (1995) or Eraker et al. (2003). When following this approach, one basically utilizes the same information, i.e. the marginal likelihoods, in order

to compare the different models but compares the marginal likelihoods directly.

As already indicated through its name, when comparing the quality of the different models' fit based on the Bayes factor we follow a Bayesian approach: Given two different models "a" and "b" and assuming positive priors for the two models, the posterior odds ratio is given as

$$\frac{\mathbb{P}(a|\lambda)}{\mathbb{P}(b|\lambda)} = \frac{\mathbb{P}(\lambda|a)}{\mathbb{P}(\lambda|b)} \cdot \frac{\mathbb{P}(a)}{\mathbb{P}(b)} = \frac{\mathcal{L}_a(\hat{\theta}_a)}{\mathcal{L}_b(\hat{\theta}_b)} \cdot \frac{\mathbb{P}(a)}{\mathbb{P}(b)} \tag{3.21}$$

where one usually refers to the first ratio of the right hand side of equation (3.21) as the Bayes factor. In other words, "posterior-odds-ratio = Bayes factor × prior-odds-ratio". The interpretation of the obtained odds ratio is the following: Given prior ignorance, i.e. $\frac{\mathbb{P}(a)}{\mathbb{P}(b)} = 1$, one would favor model "a" against "b" with a ratio of $\frac{\mathbb{P}(\lambda|a)}{\mathbb{P}(\lambda|b)}$ to be the "right" model.

In order to interpret the Bayes factor $\frac{\mathcal{L}_a(\hat{\theta}_a)}{\mathcal{L}_b(\hat{\theta}_b)}$, Kass and Raftery (1995) provide the following scale:

$$2\log\left(\frac{\mathcal{L}_a(\hat{\theta}_a)}{\mathcal{L}_b(\hat{\theta}_b)}\right) \in \begin{cases} [0,2) & \rightarrow \text{ not to mention} \\ [2,6) & \rightarrow \text{ positive evidence against "b"} \\ [6,10) & \rightarrow \text{ strong evidence against "b"} \\ [10,\infty) & \rightarrow \text{ very strong evidence against "b"} \end{cases}$$

When considering model "a" vs. "b", a Bayes factor value of 15 would for example indicate strong evidence against model "b". It is important to note that due to their marginal structure odds ratios do not automatically prefer more complex models (cf. Eraker et al. (2003)). One therefore often refers to the Bayes factor as an "automatic Occam's razor" (see Smith and Spiegelhalter (1982)).

As previously depicted, in case of Waxman Industries a jump diffusion model provides a much better fit to the data than a pure diffusion model. Not surprisingly, based on the Bayes factor and according Kass and Raftery (1995) we find very strong evidence against a BAJD and a CIR model. While in addition both models are rejected by a χ^2-significance test with p-values of approximately 0, the test does not reject the SAJD model. In case of General Electric, on the other hand, the CIR model is not rejected, and we do not find any evidence against the CIR model leading to the conclusion that the intensity dynamics of General Electric are sufficiently well described by a CIR process.

But how is the situation in the whole data set? In Table 3.7, the results for all 3241 firms are displayed. We provide the number of firms for which we find evidence against a particular model as well as the number of firms among these that

	CIR vs. BAJD	BAJD vs. SAJD	SAJD vs. SAJDM
i) Intensities based on model III (fixed coefficients)			
χ^2-significance Test (0.1%)			
Not Significant	1176 (65)	2565 (239)	2990 (376)
Significant	2065 (322)	676 (148)	251 (11)
Bayes factor			
Not to mention	1099 (39)	2078 (128)	2611 (296)
Positive	2142 (348)	1163 (259)	630 (91)
Strong	2118 (340)	879 (205)	395 (29)
Very strong	2096 (332)	703 (157)	308 (18)
ii) Intensities based on model IV (rolling estimations)			
χ^2-significance Test (0.1%)			
Not Significant	1448 (61)	2593 (229)	2951 (376)
Significant	1793 (326)	648 (158)	290 (11)
Bayes factor			
Not to mention	1365 (33)	2100 (92)	2617 (318)
Positive	1876 (354)	1141 (295)	624 (69)
Strong	1842 (339)	850 (224)	419 (26)
Very strong	1821 (336)	680 (167)	350 (17)

Table 3.7: Comparison of the single firm intensity models with respect to their statistical significance based on a standard χ^2-significance test (significance level 0.1%) and the Bayes factor with number of defaulted firms in parentheses. The total number of defaults in the considered data set is 387.

eventually defaulted (in parentheses). Furthermore, we state the number of firms (and the number of corresponding defaulters) for which a particular model was rejected at a significance level of 0.1% based on a standard χ^2-significance test. First, we observe that differences between the two data sets are minor. Second, we find for up to two thirds of the firms (56 % in case of data set (ii)) strong evidence against a CIR model when compared to a BAJD model. Even more importantly, for approximately 85% (86%) of the defaulted firms we find very strong evidence against a CIR model. Still for almost 36 % (35%) of the firms and for 67% (76%) of the defaulted firms we find evidence against a BAJD in favor of a SAJD model. Finally, Table 3.7 also provides evidence against a SAJD in case of 19% (19%) of all firms and 24 % (18%) of the defaulted firms.

In total, we see evidence for jumps of the intensities, in particular, for firms of bad creditworthiness which eventually defaulted. We therefore conclude that jumps should be taken into consideration when setting up a model of the default intensities for firms of lower creditworthiness. Also, in this case it is advisable to incorporate a self-affecting feature into a model, while the memory effect seems to be of lower relevance.

To the best of our knowledge, we are the first who investigate the dynamics of intensities under the objective, i.e. real-world, measure. Our findings are of great importance from a credit portfolio management perspective and for model specification under the real-world measure in general. So far, Mortensen (2006) has pointed out the importance of jumps under the risk-neutral measure for being able to calibrate an intensity-based credit portfolio model to quoted prices of standardized structured credit products such as the Itraxx tranches. However, due to manifold possibilities of switching from the risk-neutral to the actual measure, in particular, when a risk premium is attached to the default event itself it is important to note that the results of Mortensen (2006) do not necessarily imply jumps of the intensities under the actual measure.

Default Prediction

In Subsection 3.1.3 we used the accuracy ratio AR to measure the ability of the different models to rank firms according to their default probability. We found that ranking firms based on their current default intensity yields excellent results, in particular compared with a previous study of Duffie et al. (2007). Nevertheless, when ordering firms according to their default intensity one eventually "throws away" any time series information on the intensities. For instance, a currently high default intensity might again go down and therefore does not necessarily imply a high default likeliness of the firm in the future. The probability for default intensity changes depends on the model which we belief to be the true underlying data generating model and can be estimated from observed default intensities based on this model.

In the following, we investigate the predictive power of the different default intensity models introduced in Section 3.2. Based on parameter estimates for our four basic single firm intensity models (CIR, BAJD, SAJD, SAJDM), we calculate model-implied default probabilities and rank the firms according to this default probability. As previously mentioned, the default probability over the prediction horizon represents the natural ranking criterion when analyzing a model's predictive power. In the exponential-affine setup considered here, default probabilities can conveniently be calculated (cf. Proposition 3.2.1). Of course, when computing out-of-sample accuracy ratios parameter estimates are only based on past intensity

observations.

	1985-2005		1993-2005	
	1-year horizon	5-year horizon	1-year horizon	5-year horizon
In-sample				
Current Intensity	0.908	0.797	0.920	0.796
CIR Model	0.909	0.787	0.919	0.773
BAJD Model	0.897	0.861	0.912	0.869
SAJD Model	0.911	0.866	0.925	0.874
SAJDM Model	0.914	0.870	0.930	0.880
Out-of-sample				
Current Intensity	0.908	0.797	0.920	0.796
CIR Model	0.907	0.792	0.919	0.793
BAJD Model	0.894	0.768	0.903	0.755
SAJD Model	0.898	0.771	0.906	0.759
SAJDM Model	0.897	0.772	0.906	0.762

Table 3.8: Accuracy ratios of the single firm intensity models averaged from 1985 on and from 1993 on. Underlying default intensities have been estimated based on model IV with rolling, five-year estimation windows. The total number of defaults in the considered data set was 387.

In Table 3.8, one-year ahead and five-year ahead accuracy ratios – averaged from 1985 and from 2003 on – are displayed for the considered models. Moreover, we provide accuracy ratios derived from ranking firms merely based on their current intensity. The underlying default intensities have been estimated based on model IV with rolling estimations.[11] We observe that accuracy ratios based on current intensities are higher than those reported in Table 3.2. This has to be attributed to the definition of the accuracy ratio that we use. Recall that our definition follows Duffie et al. (2007); we do not normalize the accuracy ratio by the fraction of survivors as suggested by Tasche (2007). We already pointed out that low default rates can favor high accuracy ratios in this case.[12]

In addition, we find that out-of-sample accuracy ratios for the intensity models are

[11]Results remain qualitatively the same for default intensities estimated with model III and fixed coefficients. Only the levels of the accuracy ratios are generally lower since in this case the rating is not included as an explanatory variable in the intensity estimation, see Subsection 3.1.3.

[12]The total number of defaults in the data set considered here is 387 compared with 490 in Subsection 3.1.3.

lower than the ones derived from ranking firms only based on their current intensity. Based on our results, it is advisable to simply use the current intensity for default prediction instead of ranking firms based on model-implied default probabilities. A possible explanation are estimation errors. Note that we estimate parameters for a firm the first time when at least 24 intensity observations are available particularly exposing the first estimates to estimation errors.[13]

In-sample, the jump diffusion models (BAJD, SAJD, SAJDM) do a much better job than the CIR model. In this case, estimation errors should also be significantly reduced since the parameters of each firm are only estimated once – based on the whole intensity path – and not repeatedly. Furthermore, as documented by Table 3.7 a jump-diffusion behavior of intensities can be particularly often observed in connection with defaults. In-sample, parameter estimates therefore reflect the later occurrence of the jumps (and defaults). Nevertheless, the CIR model is not able to "use" this information.

In summary, the presented results provide evidence for jumps of default intensities under the real-world measure. Jumps play a particularly important role for defaults: The intensities of many firms that eventually defaulted show a jump-diffusion behavior. Nevertheless out-of-sample we cannot benefit from these findings. Accuracy ratios based on the intensity models are even worse when compared with accuracy ratios derived from current intensities. Apart from estimation errors, the problem is the following: Before a jump has not occurred, its probability is estimated to be 0, but soon after the first jumps the firm will probably be defaulted. A possible strategy to solve this problem would be to group firms and estimate probabilities for intensity jumps based on the groups, reducing estimation errors and attaching positive jump probabilities to firms which have not yet jumped. However, it is questionable if this would really improve results. Note that, for example, the inclusion of different macroeconomic variables in Section 3.1 had no effect on the models' predictive power because it corresponded to a shift of all default intensities (probabilities). Here, we would shift the default probabilities of possibly large groups by attaching a particular jump probability.

3.4 Assessing Model Risk: Conditional Independence vs. Contagion

In Subsection 2.6.2, we have discussed from a theoretical standpoint differences between a conditional independence setup and a model specification where past

[13]The issue of estimation errors will also be addressed in the next section. As we show there, estimation errors are indeed considerable even when the intensities can be much more frequently observed than it is the case here.

defaults are allowed to influence the default probabilities of the other firms in the portfolio. In Section 3.1, we showed that the doubly stochastic assumption is not rejected in the context of actual corporate bond default data when default intensities are estimated out-of-sample with five-year estimation windows. However, as pointed out by Lando and Nielsen (2008) the statistical tests conducted in Section 3.1 are unfit to uncover a special type of contagion in which the default of one firm affects the factor variables and thus the default intensity of others. The results of Section 3.1 have rather to be interpreted as follows: Observed defaults are likely to have been generated by the aggregated path of estimated default intensities and we need not introduce contagion effects *in addition* to the already used explanatory variables to explain the observed default clustering.

Therefore, there still remains the risk that one works with a model that is based on the conditional independence assumption while in reality contagion effects account to a possibly large extent for observed defaults. In the following, we want to quantify this risk and focus on the following questions:

- *Given that the true data generating process exhibits contagion effects, what are the implications of calibrating a conditional independence model to this data?*

- *Would we incur considerable errors in forecasting the portfolio loss distribution in this case?*

- *Would we incur considerable errors in assessing the risk of synthetic structured credit products?*

To answer these questions, we will draw paths of the portfolio loss process and the corresponding portfolio intensity based on a model in which past defaults affect the default intensities of the firms. Afterwards, we estimate wrong models – all based on the conditional independence assumption – as well as the original model, which has actually generated the data, and investigate the ability of the models to forecast the portfolio loss distribution. Furthermore, we analyze the ability of the different models to forecast important risk measures such as the expected loss (EL) of structured credit products referencing the portfolio.

The remainder of the section is structured as follows: After a brief introduction to synthetic structured credit products and their valuation in Subsection 3.4.1, we describe the methodology in Subsection 3.4.2. Subsection 3.4.3 finally presents the results.

3.4.1 Synthetic Structured Credit Products and their Valuation

As already mentioned at the beginning of this thesis, in structured finance transactions portfolio cash flows are repackaged into new securities that are called tranches and differ in their seniority.

In the following, we restrict ourselves to synthetic structured credit products that represent pure options on the aggregated loss process

$$L(t) = \frac{1}{I} \sum_{i=1}^{I} N_i(t)$$

of the underlying credit portfolio.[14] If not stated otherwise, they can be regarded as a collection of plain-vanilla options all having as underlying the portfolio loss process L. For example, the percentage time t loss, $L^{K_l,K_u}(t)$, of a synthetic CDO tranche with *attachment* point K_l and *detachment* point K_u depends on the underlying credit portfolio loss $L(t)$ in the following way:

$$\begin{aligned} L^{K_l,K_u}(t) &:= \frac{L^{K_l,1}(t) - L^{K_u,1}(t)}{K_u - K_l} \qquad \text{with} \\ L^{K,1}(t) &= (L(t) - K)^+, \end{aligned}$$

which demonstrates the well-known option-like character of CDO tranches and illustrates that the key ingredient to assess the risk of synthetic CDO tranches is the loss distribution of the reference portfolio.

The expected loss (EL) of such a tranche is obtained by integrating over the distribution of $L(t)$.[15] As pointed out in Subsection 2.7.2, in case of a conditional independence setup using the limiting loss distribution and assuming a simple model for the default-trigger process Λ even analytical formulas can be stated for the expected loss of a CDO tranche.

The valuation of CDO tranches is not in the focus of this thesis. Nevertheless, it is worth mentioning that in this case the tranche's expected loss is the main ingredient entering into the valuation formula, too, but the expected loss is then calculated under a pricing measure. Generally, valuation of a synthetic CDO tranche is closely

[14] The word "synthetic" refers to the fact that these transactions reference portfolios of Credit Default Swaps (CDSs), i.e. synthetic portfolios. Examples are standardized contracts such as the iTraxx or CDX indices. Also, note that different to Chapter 2 we will consider in the following the normalized loss process, that is $L(t) \in [0, 1]$ for each t.

[15] A common measure for the risk of a credit-sensitive security is its expected loss. For instance, Moody's claims consistency of its structured finance ratings with the expected loss (see e.g. Yoshizawa (2003)).

related to the pricing of an ordinary CDS or a term-life insurance contract: In each case the contract's payment stream can be decomposed into a premium leg (accounting for the premia payments of the contract) and a default leg (accounting for payments with respect to defaults); the tranche spread is then determined such that the expected values of both are the same under the pricing measure at the valuation date. The only difference is that the expected loss of a CDO tranche depends on a whole portfolio and not merely on one single entity.

3.4.2 Methodology

In order to investigate the issue of model risk for the time-continuous setup considered in this thesis, we follow a methodology that is inspired by Hamerle and Rösch (2005). In their study, they examine implications of misspecified copulas for the portfolio loss distribution in a static copula credit portfolio setting. More precisely, they simulate default paths based on a so-called *t-copula* and afterwards calibrate a *Gaussian* copula (for definitions see e.g. McNeil et al. (2005)) to the generated data. The t-copula thereby represents the more complex model, while the Gaussian copula is common choice in practice. They find that parameter estimates for the Gaussian copula are biased, but that forecasts of the portfolio loss distribution may still be adequate compared with the true underlying data generating t-copula model. All in all, their findings reduce model risk regarding the copula choice in practical applications.

In the following, we conduct a similar analysis and investigate the issue of model risk with respect to the conditional independence assumption. We generate default data using a model in which this assumption is violated by allowing past defaults to affect the default intensities of the other firms; afterwards, we estimate different models, all based on the assumption of conditionally independent defaults, as well as the original model, which has actually generated the data, and finally analyze the models' ability to forecast the portfolio loss distribution.

In Subsection 2.6.2, we already introduced a model with a feedback event, i.e. a model where past defaults affect the survival probabilities of the other firms. Here, we will consider a similar model. We assume that simultaneous defaults are ruled out and that the intensity of each object in our stylized portfolio is given as

$$d\lambda_i(t) = \kappa(\eta - \lambda_i(t))dt + \sigma\sqrt{\lambda_i(t)}dW(t) + dJ(t) \qquad \lambda_i(0) = \overline{\lambda}. \tag{3.22}$$

The parameters κ, b and σ and the Brownian motion W are the same for each object i. J denotes a point process that jumps whenever L jumps and whose jump sizes are independently, $Exp\left(\frac{1}{\zeta}\right)$-distributed. Then, we have that $\lambda_i(t) = \lambda_j(t) := \lambda(t)$

for all $i \neq j$ and that

$$L(t) - \int_0^t \lambda(s) \sum_{i=1}^{I} 1_{\tau_i > s}(s) ds \tag{3.23}$$

is an $\mathbf{F}$-martingale. It is worth mentioning that without the jumps of J at defaults the intensity λ reduces to a CIR process (cf. equation (3.11)) and the portfolio model becomes a conditional independence setup. Our specification of the default intensities entails that we consider a homogeneous credit portfolio where all firms have identical default probabilities and where the risk stemming from the intensities is purely systemic; all firms are exposed to one common risk driver which consists of two factors: A diffusion component representing regular fluctuations of default rates and an independent contagion component reflecting the influence of past defaults on intensities.

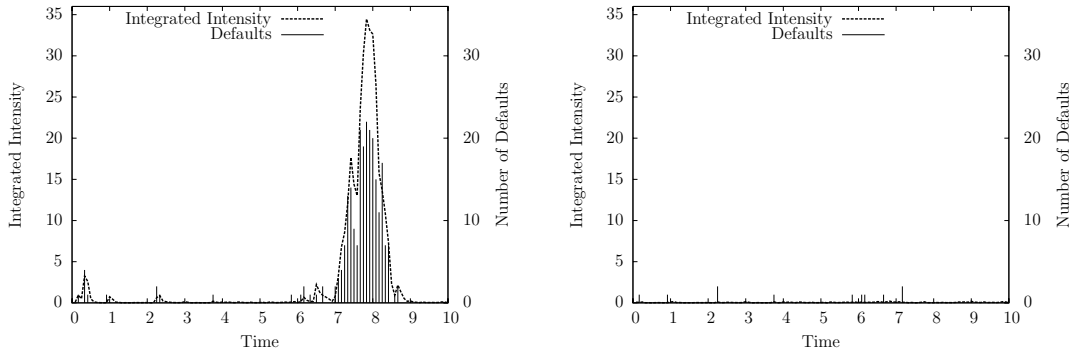


Figure 3.9: Simulated, monthly integrated portfolio intensities and defaults over a ten years period of two models that differ only in the contagion component. In both models the diffusion component is given by the mean reversion level $\eta = 0.0005$, the speed of mean reversion $\kappa = 15$ and the volatility $\sigma = 0.1$, but the jump size mean is $\frac{1}{\zeta} = 0.0138$ in the model with contagion (left) and $\frac{1}{\zeta} = 0.0$ in the model without contagion (right). The number of firms in the hypothetical portfolio is 1500.

In Figure 3.9, simulated portfolio intensities and defaults of a model with and without contagion are displayed. In the model with contagion, defaults trigger intensity jumps and possibly lead to further defaults. If no further defaults occur, the portfolio intensity will be almost deterministically pulled down again towards the mean reversion level, which is roughly $1500 \cdot 0.0005 = 0.75$ for the parametrization considered in the example (note that the graphs show the monthly, integrated portfolio intensity which is therefore pulled down towards $\approx 0.75/12 = 0.0625$). Also, for the considered parametrization most of the portfolio risk stems from the contagion component; without contagion effects the integrated intensity stays close

to 0.0625 over time which has to be attributed to the high κ value. Consequently, defaults are almost evenly distributed over time.

Estimation of the Data Generating Contagion Model

The estimation of the contagion model of equations (3.22) and (3.23) is similar to the estimation of the conditional independence models described in Subsection 3.3.1. Since the transition density of the intensities is again complex, we approximate it based on a time-discretization of the SDE (3.22). In case of the contagion model, however, we know that jumps of the intensities can only occur at defaults. The approximate transition density we therefore have to use is

$$\begin{aligned} {}_{t+\Delta,t}\tilde{g}_\lambda\left(\lambda(t+\Delta);\theta\right) = {} & 1_{L(t+\Delta)-L(t)=1}\left(1-e^{-\Delta\lambda(t)(I-L(t))}\right)h(o^{(1)},o^{(2)}) \\ & +1_{L(t+\Delta)-L(t)=0}e^{-\Delta\lambda(t)(I-L(t))}f_{N_{\kappa(\eta-\lambda(t))\Delta,\sigma\sqrt{\lambda(t)\Delta}}}\left(\lambda(t+\Delta)-\lambda(t)\right). \end{aligned}$$

Here, $f_{N_{\epsilon_1,\epsilon_2}}$ denotes the density of a Gaussian random variable with mean ϵ_1 and standard deviation ϵ_2; the function h, and the quantities $o^{(1)}$ and $o^{(2)}$ have been defined in equation (3.19). Based on this transition density, we maximize the likelihood function (3.15) with respect to the four-dimensional parameter vector $\theta = (\eta, \kappa, \sigma, \zeta)^T$.

Estimation of the Conditional Independence Models

Under the conditional independence assumption, the portfolio intensity path evolves independently of defaults and completely characterizes the portfolio loss distribution (cf. Section 2.6). Someone assuming conditional independence would therefore directly specify a model for the aggregated intensity path and totally disregard observed defaults.

In the following, we purposely act as someone who trusts in the conditional independence assumption and directly calibrate our four basic single intensity models (CIR, BAJD, SAJD, SAJDM) introduced in Section 3.2 to the observed portfolio intensity path. More precisely, we presume that the observed portfolio intensity paths $\lambda^P(t,\omega) = \sum_{i=1}^I 1_{\tau_i(\omega)>t}\lambda_i(t,\omega)$ are generated by a factor process λ^c such that

$$\lambda^P(t) = \sum_{i=1}^I 1_{\tau_i>t}\lambda_i(t) = \lambda^c(t)\sum_{i=1}^I 1_{\tau_i>t}, \tag{3.24}$$

where λ^c is a coordinate of the background process X^1, portfolio defaults as always do not affect X^1 and λ^c evolves according to one of our four basic single intensity models. For instance, such a portfolio model would result from the assumption that the default intensity λ_i of each portfolio object is $\lambda_i(t) = \lambda^c(t)$. The parameters of λ^c are then estimated based on the ML estimation procedure introduced in

Subsection 3.3.1 with the approximation of the transition density described there.

It is self-evident that we will incur an error by proceeding in this way if the data generating model does not satisfy the conditional independence assumption. To see this, consider the following simple time-discrete two-period example: In the first period a default occurs with probability 20%. Having observed a default in the first period, the default probability in the second period increases to 40% and remains at 20% otherwise. On the other hand, let us consider a model, where the default probability has the same dynamics but is independent of defaults: In the first period it is always 20% and in the second period it increases with probability 20% to 40% and remains at 20% otherwise. In both models the default probability has the same "dynamics" and the expected number of defaults is 0.44, but in the first model the probability of observing a default in both periods is $0.2 \cdot 0.4 = 0.08$ compared to only $0.048 = 0.2 \cdot 0.2 \cdot 0.4 + 0.2 \cdot 0.8 \cdot 0.2$ in the second model. Since – roughly speaking – default probabilities in the time-discrete setting correspond to default intensities in the time-continuous setup, our example shows that even if we are able to perfectly describe the dynamics of the intensities we will incur an error and underestimate the tail probabilities of the loss distribution in case that the conditional independence assumption is violated.[16] Nonetheless, knowing that we make an error, we want to examine its size and character.

Estimation based on the Loss Process

Model estimation solely based on the aggregated intensity process eventually "throws away" any information on defaults, which is justified in case that the data generating process obeys the conditional independence assumption. The approach is feasible as long as information on the portfolio intensity is available. However, in reality the portfolio intensity is usually unobserved; only defaults can be observed. The intensity can then be estimated based on observed defaults (cf. Section 3.1), but depending on the respective application and the available data this can possibly become a very involved task. Therefore, in many applications one would have to directly estimate the models based on observed defaults or at least take these defaults into account, too.

Furthermore, estimators based on the portfolio intensity and on the loss realizations should – disregarding estimation errors – yield the same parameter values in a conditional independence setup. Consequently, even if the portfolio intensity is observed it is advisable to conduct an estimation based on observed defaults since parameter estimates derived from actual defaults represent a good robustness check

[16] Of course, presuming that the dependence between intensity and loss process is positive. Otherwise, we would overestimate the tail probabilities.

for intensity-based estimates.

Given K independent realizations $l_1, \ldots, l_K$ of the portfolio loss, the corresponding likelihood is given as

$$\mathcal{L}(\theta; l) = \prod_{k=1}^{K} g_L(l_k; \theta)$$

where $f_L(x; \theta)$ denotes the probability distribution function of the portfolio loss; for the computation of the portfolio loss distribution in a conditional independence setup we refer to Subsection 2.7.2. To derive the distribution for the introduced contagion model, we have to rely on Monte-Carlo methods.[17] The length of the parameter vector θ depends on which of the models we choose for the portfolio intensity. Estimation is conducted via ML.

3.4.3 Estimation Results and the Quantification of Model Risk

By introducing estimation procedures for the portfolio models considered in this section, we finally set the stage to answer the question which is at the heart of this section: What are the implications of estimating conditional independence models based on default data which has actually been generated by a contagion driven model?

Using Algorithm A.0.2 in Appendix A, we simulate 1000 paths of default history (defaults and aggregated portfolio intensity) with length ten years on an equidistant time grid of size $\frac{1}{3600}$. This corresponds to observing portfolio intensity and defaults on a daily basis. We examine six different stylized portfolios: Each portfolio consists of 1500 firms all having default intensities as specified in equation (3.22) but portfolios differ in intensity parametrizations: We consider two portfolios (I and II) with a low ten-year default probability of roughly 1%, two portfolios (III and IV) with a medium default probability of about 3-6% and two portfolios (V and VI) showing a high ten-year default probability of roughly 16%. Furthermore, in portfolios I, III and V the portfolio intensity shows a diffusion-like behavior, i.e. jumps are smaller and the influence of the Brownian component is more pronounced, while in portfolios II, IV and VI jumps account for most of the variation in the portfolio intensity. For instance, the path displayed in Figure 3.9 is a representative of portfolio VI. In each portfolio, contagion effects significantly contribute to the default probability of the firms meaning that without these effects the default probability of the firms would be reduced by at least 50%.

[17] In the estimation procedure, we simulated 50000 paths of default history for each distribution.

Estimation Results

In Table 3.9 on p. 130, average parameter estimates based on the portfolio intensity for three conditional independence models (CIR, BAJD, SAJD) as well as for the original contagion model that has generated the data (Contagion Model) are presented.[18] The true parameters of the data generating model are stated, too (True Contagion Model). Not surprisingly, the models with jumps provide a better fit to the data, particularly in case of portfolios II, IV and VI where jumps of the intensity account for most of its volatility.

As usual, we find that the volatility σ of the Brownian component can be identified best. Only for the CIR model, parameter estimates are biased because in this case the diffusion component has to account for the whole volatility generated by the underlying contagion model, i.e. for fluctuations due to the Brownian motion *and* due to jumps. However, apart from σ standard deviations of the estimators are considerable. Specifically, in case of low or medium risk portfolios with a diffusion-like character of the portfolio intensity (portfolios I and III), the mean reversion level η and the jump parameters μ, ζ and $\xi^{(1)}$ show high variances for all models. First, this has to be attributed to the fact that there are fewer jumps for low risk portfolios, which naturally increases the variance of the estimators.[19] Second, jumps are smaller for diffusion-like specifications of the portfolio intensity, and it is therefore much harder to separate them from the diffusion part. As jumps become more pronounced, estimation errors decrease. The high values for $\xi^{(1)}$ in case of the SAJD model can be explained by the number of portfolio objects considered: The probability of observing an intensity jump over the next infinitesimally small time step Δ, is $\Delta \cdot \sum_{i=1}^{I} 1_{\tau_i > t} \lambda(t)$ in the true model, where $\sum_{i=1}^{I} 1_{\tau_i > t}$ will be most of the time close to 1500 for the low and medium risk portfolios. For portfolios I, II, III and V we obtain an $\xi^{(1)} > 1500$, but note that in these cases the estimated mean of the jump size distribution is also smaller than the true value.

All in all, we find that for all models estimation errors are considerable and errors decrease as the jump part becomes more dominant. Nonetheless, at this stage it is hard to tell whether we would overstate or understate the risk of the underlying portfolios using the derived estimates.

As a robustness check, we also conduct an estimation based on the 1000 portfolio loss realizations that we have given. The results of this estimation can be found

[18] The table does not contain estimates for the SAJDM model introduced in Section 3.2, since variances of the estimators were high for this model and the model did not provide a better fit to the data than the other considered conditional independence models.

[19] Remember that in the data generating model jumps occur only at defaults. Naturally, there are therefore fewer jumps for low and medium risk model specifications.

in Table 3.11 on p. 132.[20] Since in such a static application of the model the parameters of the portfolio intensity are difficult to identify separately, we merely estimate two parameters of each model and keep the remaining ones fixed at the values reported in Table 3.9.[21] For the CIR model we calculate the mean reversion level η and the volatility σ, in case of the jump-diffusion models we estimate the jump parameters and for the contagion model we calibrate the mean reversion level η and the jump size mean ζ. For the CIR model, we find significant deviations of the estimates from the portfolio intensity-based ones. Since in this case the new parameters imply a by far more risky parametrization (particularly for portfolio IV), this already indicates that the portfolio intensity-based estimates of the CIR model will tend to underestimate the true portfolio risk.

Forecasting the Loss Distribution and Assessing the Risk of Structured Credit Products

Table 3.10 on p. 131 summarizes results for the average parameter estimates presented in Table 3.9, which were based on the portfolio intensity. It shows descriptive statistics of ten-year loss distributions.[22] Also, the expected loss (EL) of the whole portfolio as well as of hypothetical CDO tranches referencing the portfolio (cf. Subsection 3.4.1) are presented.[23] As is common in the credit risk literature, in all computations we presumed a constant Loss-Given Default (LGD) rate of 60%. In addition, the table contains results for a model whose marginal dynamics of the portfolio intensity are exactly the same as in the underlying contagion model but whose intensity evolves independently of defaults (True Intensity Model without Contagion). The corresponding loss distribution would be obtained if we were able to correctly model and estimate the marginal dynamics of the portfolio intensity but wrongly assume conditional independence. In Subsection 3.4.2, we pointed out that this inevitably leads to an underestimation of the loss distribution's tail probabilities. However, as Table 3.10 on p. 131 shows this underestimation is not very large.

In addition, we find that the BAJD and SAJD model tend to overestimate the

[20] For ease of exposition, we only display the results for portfolios III and IV. For the other portfolios, the results are qualitatively the same.

[21] From a static point of view, only the distributional properties of the integrated portfolio intensity are relevant. However, different parameters have a similar effect on this distribution. For example, the speed of mean reversion κ, the volatility σ, the jump intensity parameters μ, $\xi^{(1)}$ and the jump size mean ζ jointly control its variance.

[22] For the models where we had to rely on Monte-Carlo simulations in order to compute the portfolio loss distribution we drew 100000 random samples for each distribution.

[23] The considered attachment and detachment points of the tranches are common. For instance, standardized CDO tranches referencing the iTraxx index exhibit the same attachment and detachment points.

expected portfolio loss and therefore the default probability of the portfolio objects. Contrarily, the CIR model underestimates the expected loss for jump-like specifications of the portfolio intensity (portfolios II, IV and VI). This means that – although parameter estimates for the CIR model were biased in this case– the model is not capable of explaining portfolio defaults when jumps account for most of the portfolio intensity's volatility.

Particularly for jump-like specifications of the portfolio intensity, the CIR model also strongly understates the tail probabilities of the loss distribution. The same holds true for the BAJD model. Apart from portfolio I, the 99.9% quantile of the loss distribution computed with a CIR or BAJD model is always smaller than the true quantile. Contrarily, for portfolios I, III and VI the quantiles obtained with the SAJD model are larger than those of the true model. For the other portfolios, the SAJD model underestimates the risk but is closer to the true values than the CIR or BAJD model. This can be attributed to the self-affecting model feature: Jumps of the portfolio intensity make further jumps and thus defaults more likely meaning that in this case more probability mass is shifted into the tails of the portfolio loss distribution. Note that in the BAJD model the jump likelihood of the intensity is constant and intensity jumps are evenly distributed over time (cf. Figure 3.7 on p. 108). For portfolios III and V, quantiles computed with the SAJD model are even closer to the true values than those obtained from the estimated contagion model (Contagion Model). When comparing results from the estimated contagion model with those of the True Intensity Model without Contagion, we find that calculated quantiles based on the latter are closer to the true values for portfolios I, III, IV, V and VI.

Results remain qualitatively the same when considering the expected loss of CDO tranches with the different portfolios as underlying: CIR and BAJD model would lead to an understatement of risk, while risk figures obtained from the SAJD model are slightly more conservative than the true values. Again, the estimated SAJD and contagion model lead to comparable results. In particular, using a SAJD model one would not have understated the risk of the tranches more often than with the contagion model.

In Table 3.12 on p. 132, we present results which are computed with the parameter estimates of Table 3.11. Recall that these estimates were based on the portfolio loss. We find that – aside from the BAJD model in case of portfolio IV – all conditional independence models now tend to overstate the tail probabilities of the portfolio loss distribution and the risk of the more senior tranches; the estimated default probabilities are close to the true ones. This shows that all conditional independence models are able to model the loss distribution if they are estimated based on observed losses.

	Average η	Average κ	Average σ	Average μ	Average ζ	Average $\xi^{(1)}$	Average Log-Likelihood
Portfolio I (1.08% Default Probability, Diffusion-like behavior of the portfolio intensity)							
CIR Model	0.00066 (0.00098)	0.85860 (0.63916)	0.05683 (0.00281)	–	–	–	30643
BAJD Model	0.00010 (0.00020)	1.96765 (1.06293)	0.04893 (0.00127)	84.89080 (97.98832)	0.00017 (0.00024)	–	31285
SAJD Model	0.00033 (0.00068)	1.45352 (0.66740)	0.05162 (0.00142)	–	0.00022 (0.00019)	7944.61909 (3094.73559)	31059
Contagion Model	0.00026 (0.00024)	1.36746 (0.56829)	0.05495 (0.00315)	–	0.00091 (0.00368)	–	30754
True Contagion Model	0.00040	1.50000	0.05000	–	0.00070	–	–
Portfolio II (1.08% Default Probability, Jump-like behavior of the portfolio intensity)							
CIR Model	0.00059 (0.00006)	13.62045 (1.70017)	0.10184 (0.00262)	–	–	–	26919
BAJD Model	0.00054 (0.00005)	14.55317 (1.29745)	0.09833 (0.00139)	2.35634 (7.30635)	0.00419 (0.00172)	–	27340
SAJD Model	0.00054 (0.00004)	14.93498 (1.30244)	0.09857 (0.00127)	–	0.00408 (0.00153)	1814.53303 (837.33557)	27346
Contagion Model	0.00055 (0.00004)	14.90333 (1.30273)	0.09911 (0.00128)	–	0.00471 (0.00144)	–	27325
True Contagion Model	0.00055	15.00000	0.10000	–	0.00500	–	–
Portfolio III (6.10% Default Probability, Diffusion-like behavior of the portfolio intensity)							
CIR Model	0.00839 (0.00814)	0.82657 (0.57050)	0.10590 (0.00202)	–	–	–	24040
BAJD Model	0.00304 (0.00272)	1.39846 (0.69509)	0.09961 (0.00171)	12.06576 (23.52674)	0.00066 (0.00032)	–	24147
SAJD Model	0.00183 (0.00135)	2.57481 (0.86002)	0.09860 (0.00175)	–	0.00050 (0.00019)	5139.89561 (3293.43322)	24141
Contagion Model	0.00266 (0.00143)	1.75822 (0.56590)	0.10093 (0.00172)	–	0.00069 (0.00010)	–	23815
True Contagion Model	0.00260	1.50000	0.10000	–	0.00070	–	–
Portfolio IV (3.9% Default Probability, Jump-like behavior of the portfolio intensity)							
CIR Model	0.00064 (0.00019)	11.10342 (3.01584)	0.11701 (0.02355)	–	–	–	25536
BAJD Model	0.00049 (0.00005)	14.16444 (1.00765)	0.09894 (0.00132)	5.10384 (7.69370)	0.00981 (0.00309)	–	26660
SAJD Model	0.00049 (0.00004)	14.82276 (0.91136)	0.09888 (0.00125)	–	0.00866 (0.00229)	1470.15231 (519.06702)	26716
Contagion Model	0.00050 (0.00005)	14.83484 (0.91685)	0.09929 (0.00129)	–	0.00910 (0.00194)	–	26665
True Contagion Model	0.00050	15.00000	0.10000	–	0.01000	–	–
Portfolio V (16.07% Default Probability, Diffusion-like behavior of the portfolio intensity)							
CIR Model	0.02020 (0.00776)	1.02429 (0.52463)	0.10570 (0.00162)	–	–	–	21190
BAJD Model	0.00868 (0.00389)	1.26446 (0.54975)	0.10006 (0.00165)	17.83564 (13.24354)	0.00080 (0.00020)	–	21238
SAJD Model	0.00742 (0.00298)	2.43960 (0.75957)	0.09799 (0.00191)	–	0.00061 (0.00014)	3103.41721 (2037.05905)	21241
Contagion Model	0.00935 (0.00280)	1.88721 (0.51283)	0.09972 (0.00117)	–	0.00070 (0.00006)	–	20500
True Contagion Model	0.00800	1.50000	0.10000	–	0.00070	–	–
Portfolio VI (16.2% Default Probability, Jump-like behavior of the portfolio intensity)							
CIR Model	0.00181 (0.00136)	5.25416 (4.35906)	0.21798 (0.08466)	–	–	–	22076
BAJD Model	0.00050 (0.00005)	13.60518 (0.83200)	0.10028 (0.00170)	18.64582 (12.83383)	0.01578 (0.00361)	–	24523
SAJD Model	0.00049 (0.00005)	14.74777 (0.59349)	0.09872 (0.00133)	–	0.01283 (0.00245)	1321.02317 (340.10108)	24755
Contagion Model	0.00050 (0.00005)	14.78073 (0.60041)	0.09899 (0.00135)	–	0.01298 (0.00207)	–	24589
True Contagion Model	0.00050	15.00000	0.10000	–	0.01380	–	–

Table 3.9: Average parameter estimates based on the aggregated portfolio intensity and 1000 simulated paths of default history. Estimates are presented for three models (CIR, BAJD, SAJD) satisfying the conditional independence assumption as well as for estimates of the model that has generated the data (Contagion Model). Six different portfolios are considered: Low risk portfolios (I and II), medium risk portfolios (III and IV) and high risk portfolios (V and VI). In portfolios I, III and V the portfolio intensity shows a diffusion-like behavior; in portfolios II, IV and VI it is strongly driven by jumps. Standard deviations in parentheses.

	EL of CDO Tranches $[K_l, K_u]$ and of the whole Portfolio							Quantiles of the Loss Distribution			
	[0,0.03]	[0.03,0.06]	[0.06,0.09]	[0.09,0.12]	[0.12,0.22]	[0.22,1]	Portfolio	0.95	0.99	0.995	0.999
Portfolio I (1.08% Default Probability, Diffusion-like behavior of the portfolio intensity)											
CIR Model (CI)	0.1240	0.0000	0.0000	0.0000	0.0000	0.0000	0.0037	0.010	0.015	0.018	0.023
BAJD Model (CI)	0.9993	0.3982	0.0001	0.0000	0.0000	0.0000	0.0419	0.052	0.056	0.058	0.062
SAJD Model (CI)	0.5207	0.1798	0.0472	0.0057	0.0001	0.0000	0.0226	0.073	0.096	0.104	0.120
Contagion Model	0.2541	0.0587	0.0183	0.0048	0.0004	0.0000	0.0101	0.048	0.088	0.103	0.135
True Contagion Model	0.2102	0.0044	0.0001	0.0000	0.0000	0.0000	0.0064	0.021	0.034	0.040	0.054
True Intensity Model without Contagion (CI)	0.2131	0.0031	0.0000	0.0000	0.0000	0.0000	0.0065	0.020	0.032	0.037	0.050
Portfolio II (1.08% Default Probability, Jump-like behavior of the portfolio intensity)											
CIR Model (CI)	0.1163	0.0000	0.0000	0.0000	0.0000	0.0000	0.0035	0.006	0.007	0.007	0.008
BAJD Model (CI)	0.2411	0.0000	0.0000	0.0000	0.0000	0.0000	0.0072	0.011	0.013	0.014	0.015
SAJD Model (CI)	0.2105	0.0000	0.0000	0.0000	0.0000	0.0000	0.0063	0.012	0.014	0.016	0.018
Contagion Model	0.2046	0.0000	0.0000	0.0000	0.0000	0.0000	0.0061	0.013	0.016	0.018	0.022
True Contagion Model	0.2148	0.0000	0.0000	0.0000	0.0000	0.0000	0.0064	0.014	0.018	0.020	0.024
True Intensity Model without Contagion (CI)	0.2154	0.0000	0.0000	0.0000	0.0000	0.0000	0.0065	0.012	0.015	0.016	0.020
Portfolio III (6.10% Default Probability, Diffusion-like behavior of the portfolio intensity)											
CIR Model (CI)	0.9554	0.4144	0.0607	0.0054	0.0001	0.0000	0.0431	0.076	0.095	0.103	0.122
BAJD Model (CI)	0.9947	0.5545	0.0397	0.0004	0.0000	0.0000	0.0477	0.070	0.081	0.086	0.095
SAJD Model (CI)	0.9678	0.6625	0.2822	0.0720	0.0030	0.0000	0.0598	0.110	0.132	0.140	0.155
Contagion Model	0.8281	0.2049	0.0216	0.0014	0.0000	0.0000	0.0317	0.062	0.082	0.089	0.105
True Contagion Model	0.8400	0.3004	0.0698	0.0121	0.0005	0.0000	0.0367	0.080	0.106	0.116	0.138
True Intensity Model without Contagion (CI)	0.8508	0.2998	0.0656	0.0105	0.0004	0.0000	0.0368	0.078	0.104	0.114	0.138
Portfolio IV (3.9% Default Probability, Jump-like behavior of the portfolio intensity)											
CIR Model (CI)	0.1276	0.0000	0.0000	0.0000	0.0000	0.0000	0.0038	0.006	0.008	0.008	0.009
BAJD Model (CI)	0.7647	0.0078	0.0000	0.0000	0.0000	0.0000	0.0232	0.032	0.036	0.038	0.040
SAJD Model (CI)	0.4525	0.0363	0.0021	0.0001	0.0000	0.0000	0.0147	0.039	0.056	0.063	0.078
Contagion Model	0.5014	0.0723	0.0059	0.0001	0.0000	0.0000	0.0174	0.048	0.067	0.074	0.088
True Contagion Model	0.5901	0.1595	0.0282	0.0025	0.0000	0.0000	0.0234	0.064	0.087	0.095	0.110
True Intensity Model without Contagion (CI)	0.6138	0.1528	0.0230	0.0016	0.0000	0.0000	0.0237	0.062	0.083	0.091	0.106
Portfolio V (16.07% Default Probability, Diffusion-like behavior of the portfolio intensity)											
CIR Model (CI)	1.0000	0.9988	0.8784	0.3752	0.0180	0.0000	0.0994	0.134	0.152	0.159	0.174
BAJD Model (CI)	1.0000	0.9998	0.9341	0.4579	0.0200	0.0000	0.1038	0.135	0.150	0.156	0.168
SAJD Model (CI)	1.0000	0.9996	0.9531	0.6254	0.0625	0.0000	0.1136	0.153	0.170	0.177	0.190
Contagion Model	1.0000	0.9929	0.7722	0.2533	0.0086	0.0000	0.0914	0.126	0.143	0.149	0.162
True Contagion Model	1.0000	0.9787	0.7588	0.3599	0.0355	0.0000	0.0965	0.146	0.170	0.179	0.198
True Intensity Model without Contagion (CI)	1.0000	0.9825	0.7662	0.3567	0.0327	0.0000	0.0964	0.144	0.167	0.176	0.194
Portfolio VI (16.2% Default Probability, Jump-like behavior of the portfolio intensity)											
CIR Model (CI)	0.3511	0.0000	0.0000	0.0000	0.0000	0.0000	0.0105	0.018	0.022	0.023	0.027
BAJD Model (CI)	1.0000	1.0000	0.9959	0.6444	0.0131	0.0000	0.1105	0.130	0.138	0.141	0.147
SAJD Model (CI)	0.7414	0.4610	0.3087	0.1969	0.0669	0.0007	0.0585	0.177	0.241	0.264	0.306
Contagion Model	0.8425	0.6857	0.5357	0.3506	0.0715	0.0000	0.0796	0.166	0.194	0.203	0.222
True Contagion Model	0.8730	0.7617	0.6504	0.4905	0.1382	0.0001	0.0972	0.186	0.213	0.222	0.240
True Intensity Model without Contagion (CI)	0.8876	0.7677	0.6572	0.4947	0.1359	0.0000	0.0978	0.185	0.210	0.218	0.234

Table 3.10: Ten-year portfolio loss distributions derived with average parameter estimates based on the portfolio intensity. Results are presented for estimates of four conditional independence models (CIR, BAJD, SAJD) as well as for estimates of the model that has generated the data (Contagion Model). Also, the true loss distribution is stated for each portfolio (True Contagion Model) and the loss distribution which results from a conditional independence model with the same marginal intensity dynamics like the data generating model (True Intensity Model without Contagion). Models satisfying the conditional independence assumption are marked by "CI".

	η	κ	σ	μ	ζ	$\xi^{(1)}$	Log-Likelihood
Portfolio III (6.10% Default Probability, Diffusion-like behavior of the portfolio intensity)							
CIR Model	0.00820 (0.00839)	0.82657	0.17889 (0.10590)	–	–	–	-5246
BAJD Model	0.00304	1.39846	0.09961	0.19240 (12.06576)	0.04074 (0.00066)	–	-5276
SAJD Model	0.00183	2.57481	0.09860	–	0.00290 (0.00050)	714.63401 (5139.89561)	-5260
Contagion Model	0.00200 (0.00266)	1.75822	0.10093	–	0.00099 (0.00069)	–	-5278
True Contagion Model	0.00260	1.50000	0.10000	–	0.00070	–	–
Portfolio IV (5.87% Default Probability, Jump-like behavior of the portfolio intensity)							
CIR Model	0.02354 (0.00064)	11.10342	4.90064 (0.11701)	–	–	–	-5148
BAJD Model	0.00049	14.16444	0.09894	0.17421(5.10384)	0.58323(0.00981)	–	-5108
SAJD Model	0.00049	14.82276	0.09888	–	0.01100 (0.00866)	1302.80978 (1470.15231)	-5073
Contagion Model	0.00051(0.00050)	14.83484	0.09929	–	0.00991 (0.00910)	–	-5029
True Contagion Model	0.00050	15.00000	0.10000	–	0.01000	–	–

Table 3.11: Parameter estimates based on the portfolio loss and 1000 simulated paths of default history. Estimates are presented for three models (CIR, BAJD, SAJD) satisfying the conditional independence assumption as well as for the original model that has generated the data (Contagion Model). For each model, only two parameters are estimated, while the others are kept fixed at the values reported in Table 3.9. For the estimated parameters, the corresponding values of Table 3.9 are given in parentheses.

	EL of CDO Tranches $[K_l, K_u]$ and of the whole Portfolio						Quantiles of the Loss Distribution			
	[0,0.03]	[0.03,0.06]	[0.06,0.09]	[0.09,0.12]	[0.12,0.22]	[0.22,1]	Portfolio	0.95	0.99	0.995	0.999
Portfolio III (6.10% Default Probability, Diffusion-like behavior of the portfolio intensity)											
CIR Model (CI)	0.8629	0.3481	0.1064	0.0316	0.0034	0.0000	0.0408	0.092	0.131	0.146	0.179
BAJD Model (CI)	0.8713	0.4282	0.1698	0.0573	0.0065	0.0000	0.0465	0.108	0.145	0.160	0.192
SAJD Model (CI)	0.8436	0.3442	0.1194	0.0408	0.0052	0.0000	0.0410	0.098	0.140	0.156	0.190
Contagion Model	0.8388	0.3676	0.1250	0.0343	0.0025	0.0000	0.0412	0.096	0.128	0.140	0.162
True Contagion Model	0.8407	0.3008	0.0688	0.0123	0.0005	0.0000	0.0367	0.080	0.106	0.117	0.138
Portfolio IV (3.9% Default Probability, Jump-like behavior of the portfolio intensity)											
CIR Model (CI)	0.9570	0.5359	0.1623	0.0363	0.0023	0.0000	0.0510	0.098	0.126	0.138	0.163
BAJD Model (CI)	0.6300	0.1545	0.0185	0.0012	0.0000	0.0000	0.0241	0.060	0.080	0.088	0.104
SAJD Model (CI)	0.5635	0.1452	0.0363	0.0076	0.0005	0.0000	0.0226	0.067	0.098	0.111	0.138
Contagion Model	0.6038	0.1702	0.0320	0.0032	0.0000	0.0000	0.0243	0.066	0.090	0.098	0.114
True Contagion Model	0.5901	0.1595	0.0282	0.0025	0.0000	0.0000	0.0234	0.064	0.087	0.095	0.110

Table 3.12: Ten-year portfolio loss distributions derived with parameter estimates based on the portfolio loss. Results are presented for estimates of thre conditional independence models (CIR, BAJD, SAJD) as well as for estimates of the original model that has generated the data (Contagion Model). The true loss distribution is stated for each portfolio, too (True Contagion Model). Models satisfying the conditional independence assumption are marked by "CI".

All in all, it appears that for applications the issue of estimation risk is more important than the question of whether defaults are conditionally independent or not. If we were able to perfectly model and estimate the dynamics of the portfolio intensity but wrongly assume conditional independence, we would have predicted the quantiles of the loss distribution more accurate in more than 50% of the examples than with the estimated contagion model (cf. Table 3.10). Also, our results do not suggest that the SAJD model does a worse job than the estimated contagion model. The model performance is comparable to the one of the estimated contagion model. In particular, we would not have understated the risk of structured credit products using a SAJD model. Rather, risk figures tend to be conservative in this case.

Furthermore, it is important to note that the loss distributions presented in Table 3.10 are based on *average* parameter estimates. However, in reality only one default history will be available and estimation errors can have a considerable effect on the results. For instance, in case of portfolio IV the average jump size mean estimated for the Contagion Model is 0.0091 (cf. Table 3.9) yielding a 99.9% quantile of 0.088 (cf. Table 3.10). Using instead a jump size mean of 0.011 in the computation of the loss distribution, which corresponds to roughly adding one standard deviation to the average value, and keeping the other parameters fixed at their average estimates, would result in a 99.9% quantile of 0.151.

Based on our findings, we can finally answer the questions raised at the beginning of this section: First, loss distributions computed from models satisfying the conditional independence assumption and from models with contagion differ when the data generating model exhibits contagion effects. However, differences are minor in comparison with estimation errors.[24]

In this context, it is important to note that we were able to observe simulated default intensities on a daily basis but in reality they cannot usually be observed on more than a monthly basis (cf. Subsection 3.1). Second, using a CIR or BAJD model we would have understated the tails of the portfolio loss distribution. This does not hold true for a SAJD model whose performance is comparable to that of the contagion model. Third, with a CIR or BAJD model we would have incurred considerable errors in assessing the risk of CDO tranches referencing the portfolios, while for the SAJD model risk measures tend to be rather conservative. Therefore, the choice of the right model for the intensity dynamics still remains an important issue. To reduce model risk, it is worthwhile to conduct additional estimations based on realized portfolio losses. If the resulting parameters imply a by far more risky model parametrization than those obtained from the portfolio intensity, this will be a strong indicator for a model that misses large parts of the true portfolio

[24]For static credit portfolio models, the influence of estimation errors on risk measures of credit portfolio risk has already been investigated by Löffler (2003).

risk.

In summary, our results support the continued use of the conditional independence based models introduced in Section 3.2 to model the joint dynamics of default intensities in the next section.

3.5 Portfolio Modeling

The results presented in the previous section suggest that conditional independence and contagion models lead to comparable results when applied to real data, even if the data was generated by a model in which contagion effects played a dominant role.

Our findings therefore encourage the use of the conditional independence models introduced in Section 3.2 for modeling the intensity dynamics on a portfolio level. In the following, we consider their calibration to the intensities estimated in Section 3.1 and examine implications for the modeling of structured credit products. We will try to answer the following questions:

- *Which of the different versions of the model introduced in Section 3.2 describes the portfolio intensity dynamics best?*
- *What are the implications of the model choice for the risk characteristics of structured credit products?*
- *Do simple and complex models lead to comparable results?*

We address these questions following a two step approach: First, we estimate the different model versions based on the provided default intensity data. This allows us then to compare the different model versions with respect to their statistical significance and the Bayes factors associated with them (cf. Section 3.3). Since we only observe the default intensities but not the common factor and the idiosyncratic components governing the intensities, we have to deal with unobservable processes in the estimation.

In a second step, using the estimated parameters we simulate paths of the portfolio loss process and the corresponding portfolio intensity. By doing this, we can compute a time series of ratings for hypothetical, structured credit products referencing the portfolio and eventually obtain transition matrices for these products. Finally, we compare the different models with respect to these matrices.

The remainder of the section is structured as follows: In Subsection 3.5.1, we describe the methodology. Subsection 3.5.2 contains our estimation results. In Subsection 3.5.3, transition matrices of structured credit products are derived.

3.5.1 Estimation Methodology

Figure 3.10 shows the two estimated portfolio intensity paths which could explain the observed default clustering (for details see Section 3.1).[25] Each path comprises 243 monthly observed portfolio intensities. Also, the corresponding averaged default intensities are displayed. The recessions in 1990-91 and 2001 are visible with average intensities peaking at around 0.1. As always when modeling the risk of a particular portfolio, the key challenge is to explain the dynamics of the portfolio intensity. Since we assume conditional independence, however, the portfolio intensity can directly be modeled, and defaults can be disregarded.

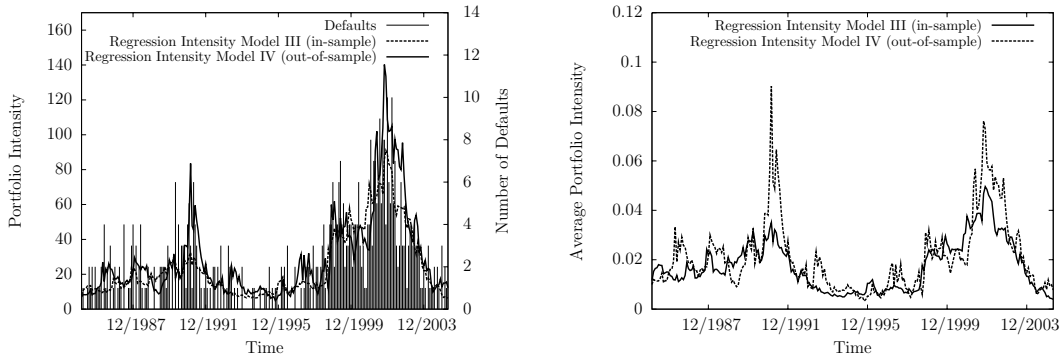


Figure 3.10: Estimated aggregated (left) and averaged (right) default intensities for the two regression intensity models that could explain the observed default clustering (for details see Section (3.1)).

Modeling the Portfolio Intensity

In order to model the portfolio intensity dynamics, we consider two basic approaches. The first approach is in line with the previous section where we presumed that the portfolio intensity λ^P is given as

$$\lambda^P(t) = \sum_{i=1}^{I} 1_{\tau_i > t} \lambda_i(t) = \lambda^c(t) \sum_{i=1}^{I} 1_{\tau_i > t}. \tag{3.25}$$

Thereby, portfolio defaults did not affect the factor process λ^c, and λ^c evolved according to one of our single firm intensity models (CIR, BAJD and SAJD model). Here, we will consider a more general specification for this factor process by assuming that λ^c is a basic jump diffusion (BJD) (see Section 3.2). As pointed out in the previous section, a model of the form given by equation (3.25) would result

[25]For ease of exposition, all results presented in the following are based on intensities that have been estimated with regression intensity model IV (out-of-sample).

from the assumption that the intensity of each portfolio object is $\lambda_i(t) = \lambda^c(t)$, i.e. that the intensities of all portfolio objects are equal. It is, however, important to note that also models with heterogeneous default intensities of the form

$$\lambda_i(t) = a_i \lambda^c(t) \tag{3.26}$$

come practically very close to the homogeneous model of equation (3.25) because it is convenient to determine the a_is in a way such that $\sum_{i=1}^{I} a_i = I$. This gives the a_is the interpretation of *portfolio weights*; they control how strong each firm contributes to the expected portfolio loss. For example, the a_is estimated in Mortensen (2006) and Feldhütter (2008) satisfy this condition. In both studies, the pricing of structured credit products in the Duffie and Gârleanu (2001) model is addressed, and CDS spreads are used as a proxy for the a_is: A firm's portfolio weight a_i is calculated as the firm's average five-year CDS spread across the estimation period divided by the average five-year CDS spread across all firms and time. We adapt this approach to our problem and compute each portfolio weight as the firm's averaged observed intensity λ_i^{Avg} divided by the intensity average λ^{Avg} in the whole sample. If the intensities of all firms were equally often observed, this would ensure that $\sum_{i=1}^{I} a_i = I$ and that the portfolio intensity is in both models the same as long as no default occurs.

Unfortunately, in our data set we do not observe the firms' intensities equally often. Also, there are defaults and other "exits" of firms from the data set, but Figure 3.11 shows that differences between the estimated factor λ^c in a homogeneous and in a heterogeneous model are still minor. Recall that the portfolio weights a_i are computed using estimated default intensities and are therefore exposed to estimation errors, too. When taking the potential estimation errors in connection with the single default intensities into account, differences between both paths in Figure 3.11 become negligible. We base our model estimation on the factor path which has been derived under the assumption of heterogeneous default intensities.

In the depicted approach, all default intensities are solely governed by one common factor λ^c and the risk stemming from the intensities is therefore purely systemic. By assuming a model for the factor process λ^c, this allows us to estimate the portfolio model more or less directly based on the observed portfolio intensity path because in this case – given the portfolio weights a_i – the factor path can immediately be derived.

In the second approach, which we describe next, this is not the case anymore. Like in the original portfolio model of Section 3.2, we now assume that the firms' default

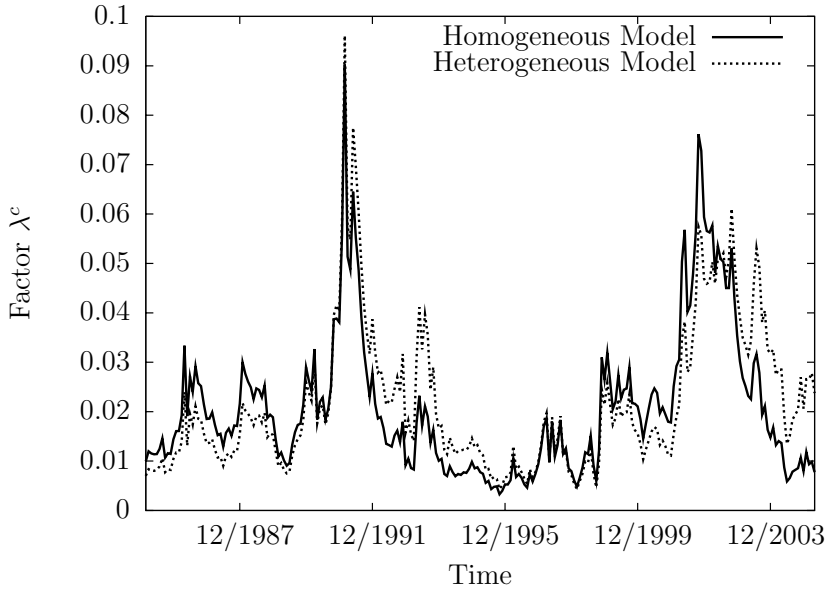


Figure 3.11: The estimated factor λ^c in a model with homogeneous intensities and in a model with heterogeneous intensities. In the homogeneous model, all default intensities are assumed to be of the form $\lambda_i = \lambda^c$. In the heterogeneous case, intensities are given as $\lambda_i = a_i \lambda^c$, where each $a_i := \frac{\lambda_i^{Avg}}{\lambda^{Avg}}$. λ_i^{Avg} and λ^{Avg} denote the average intensity of firm i and the average intensity in the whole sample, respectively.

intensities are given as

$$\lambda_i(t) = a_i \left(\lambda^c(t) + \tilde{\lambda}_i(t) \right) \tag{3.27}$$

with λ^c as always denoting the factor and $\tilde{\lambda}_i$ the idiosyncratic component of the firm's default intensity. Both processes are assumed to follow basic jump diffusion models (BJDs) as specified in equation (3.8). This implies that the portfolio intensity is given as the sum of $I + 1$ unobservable processes, which significantly complicates model inference.

To better structure the problem, we write the portfolio intensity as

$$\sum_{i=1}^{I} \lambda_i(t) = I\lambda^c(t) + \tilde{\lambda}^{idio}(t),$$

where we set $\tilde{\lambda}^{idio} := \sum_{i=1}^{I} a_i \tilde{\lambda}_i$ and assume that $\sum_{i=1}^{I} a_i = I$. In the following, we refer to $\tilde{\lambda}^{idio}$ as the idiosyncratic component of the portfolio intensity. Given homogeneity of the portfolio objects, i.e. given that $a_i = a_j = 1$ for all $i \neq j$, we can apply "calculation rule" (3.9), which yields for the idiosyncratic component of

the portfolio intensity $\tilde{\lambda}^{idio}$ that

$$\tilde{\lambda}^{idio} \quad \text{is a} \quad BJD\left(\sum_{i=1}^{I}\overline{\tilde{\lambda}}_i, \sum_{i=1}^{I}\overline{\Theta}_i, \overline{Z}, \mathcal{Q}, 1-\varpi, I\cdot\eta, \kappa, \sigma, I\cdot\mu, \zeta, \xi^{(1)}, \epsilon, \xi^{(2)}\right).$$

This means that in the parsimonious, homogeneous model specification, which we will consider in the following, the portfolio intensity is the sum of two processes. Both processes cannot still be observed, but the dimension of the problem is considerably reduced from $I+1$ to 2.

It is self-evident that by proceeding in this way we will incur an error because the underlying portfolio does definitely not satisfy the homogeneity assumption (see Figure 3.8 on p. 113 for an example of two extremely different default intensities in the data set). By assuming homogeneity, we understate the volatility of the idiosyncratic component $\tilde{\lambda}^{idio}$, i.e. the idiosyncratic component is assumed to be more diversified than it is in reality. Therefore, we can expect estimated model parameters to give more weight to the factor component because the factor component has to account for more of the portfolio intensity's overall volatility. Nevertheless, we consider our approach to be appropriate for a couple of reasons: First, it significantly simplifies the estimation procedure because the dimension of the problem is considerably reduced. Second and more important, we want to apply the estimated model to calculate transition matrices of structured credit products. The risk of these products depends on the volatility of the loss process which itself is completely determined by the volatility of the portfolio intensity. Whether the volatility of the portfolio intensity stems from the factor component or from the idiosyncratic component is not important at all as long as the volatility is sufficiently well explained.[26]

Also, this can be seen as a strong argument for the more or less direct estimation of the model in our first estimation approach of equations (3.25) and (3.26). In summary, both introduced approaches come close to top-down approaches, i.e. models that directly model the portfolio loss, but the "roots" of these approaches lie in a bottom-up model.

Estimation Methodology

Although we made simplifying assumptions to obtain a parsimonious model specification, in both approaches presented above we have to deal with unobservable processes. Direct model inference based on the ML estimation procedure introduced in Subsection 3.3.1 is only possible in the first approach when the regime

[26]The issue of the volatility "source" would become very important in a pricing context in order to compute hedge ratios for synthetic structured credit products with respect to changes in one of the underlying CDS contracts.

process Z is a trivial process, i.e. exhibits only one state. Otherwise, we have to cope with the unobserved regime process Z. In the second approach, we also have to handle the unobserved factor and idiosyncratic component that sum up to the portfolio intensity.

To tackle the problem of unobserved processes, we rely on the *expectation -maximization* (henceforth EM) algorithm, see Dempster et al. (1977); for its general examination and theoretical results see Robert and Casella (1999) and for a discussion in the context of hidden Markov models we refer to Cappé et al. (2005). Basically, the EM algorithm derives a sequence of parameter estimates $\theta^{(m)}$, which corresponds to a non-decreasing sequence of log-likelihoods. If the algorithm ever stops at a point, this point represents a local maximum of the log-likelihood (see Cappé et al. (2005), pp. 349). At the heart of the algorithm is a so-called *intermediate quantity*, which is defined as

$$\mathcal{I}(\theta, \theta') = \int \log \mathcal{L}\left(\lambda^P, w; \theta\right) p_{\mathcal{W}}(w; \theta' | \lambda^P) dw.$$

Here, $\mathcal{W}$ denotes the path of the unobserved process and w a realization. $\mathcal{L}\left(\lambda^P, w; \theta\right)$ is the complete likelihood of the observed portfolio intensity and a realization of the unobserved process w at θ. $p_{\mathcal{W}}(\cdot; \theta' | \lambda^P)$ represents the conditional density of the unobserved process given the realization of the portfolio intensity and a particular model parametrization θ'.

Having defined the intermediate quantity, we set the stage for introducing the EM algorithm:

Algorithm 3.5.1 *Set $m = 0$ and choose an initial parameter guess $\theta^{(0)}$. Proceed then as follows:*

1. *E-Step: Compute $\mathcal{I}(\theta, \theta^{(m)})$.*
2. *M-Step: Maximize the intermediate quantity $\mathcal{I}(\theta, \theta^{(m)})$ with respect to θ and set*
$$\theta^{(m+1)} = \arg\max_{\theta} \ \mathcal{I}(\theta, \theta^{(m)})$$
3. *As long as no sufficient convergence is obtained, increase m by 1 and proceed with step 1.*

The key challenge is to compute the intermediate quantity $\mathcal{I}(\theta, \theta')$ in the algorithm. In the setup considered here, it is not possible to derive this quantity analytically. Rather, we have to approximate it in the following way:

$$\mathcal{I}(\theta, \theta') \approx \tilde{\mathcal{I}}(\theta, \theta') = \frac{1}{K} \sum_{k=1}^{K} \log \mathcal{L}\left(\lambda^P, w^{(k)}; \theta\right), \tag{3.28}$$

where $w^{(k)}$ symoblizes a simulated path of the unobserved process, which is drawn from the conditional density $p_{\mathcal{W}}(\cdot;\theta'|\,\lambda^P)$. Since we rely on Monte-Carlo methods, the algorithm that we eventually apply is a so-called Monte-Carlo EM algorithm (see Robert and Casella (1999) and Cappé et al. (2005)).[27]

Depending on the form of the conditional density $p_{\mathcal{W}}(\cdot;\theta'|\,\lambda^P)$, we have to consider different simulation procedures. If direct simulation of the unobserved process from its conditional density $p_{\mathcal{W}}(\cdot;\theta'|\,\lambda^P)$ is possible, we draw i.i.d. paths. This will be the case in our second approach for models in which the regime process Z is a trivial one-state process. To sample from the conditional density we draw paths of the idiosyncratic component $\tilde{\lambda}^{idio}$ of the portfolio intensity according to the assumed model. A path $\tilde{\lambda}^{idio(k)}$ will be saved if it satisfies $\lambda^P(t_l) \geq \tilde{\lambda}^{idio}(t_l)$ at each point in time t_l. Otherwise, the path will be discarded. If the path is saved, the corresponding factor path $\lambda^{c(k)}$ is given as $\lambda^{c(k)}(t) = \frac{\lambda^P(t)-\tilde{\lambda}^{idio}(t)}{\sum_{i=1}^{I} 1_{\tau_i>t}}$.

In the models with a non-trivial regime process Z, we have to rely on a Markov-Chain Monte-Carlo (MCMC) approach for simulation. In this case, we simulate a Markov-Chain with stationary distribution $p_{\mathcal{W}}(\cdot;\theta'|\,\lambda^P)$. As described in Appendix A, we do this by applying the Gibbs sampler, see e.g. Geman and Geman (1984).

3.5.2 Estimation Results for the Portfolio Models

In Table 3.13 parameter estimates based on our first estimation approach are presented. The approach takes the heterogeneity of the underlying portfolio into account, but assumes that intensities are solely governed by the common factor λ^c. In addition, we state the corresponding estimates that are obtained under the assumption of homogeneous default intensities. As expected, differences between the heterogeneous and the homogeneous models are only slight.

In the estimation, we observe a pattern which we have already encountered in Section 3.3: As soon as the possibility of jumps is included the mean reversion level drops, the speed of mean reversion increases and volatility moves from the Brownian part into the jump part. When incorporating a self-affecting model feature, jumps are estimated to be of pure self-affecting nature since $\mu = 0$. A comparison of the different models based on the log-likelihoods and the Bayes factors as discussed in Section 3.3 yields very strong evidence against a CIR model in favor of a BAJD model and strong evidence against a BAJD model in favor of a SAJD model eventually corresponding to evidence of jumps of the portfolio intensity. Nonetheless, at this stage it is hard to tell whether the estimated models really imply different

[27] For implementation details regarding the Monte-Carlo EM algorithm we refer to Appendix A.

	CIR Model	BAJD Model	SAJD Model
Mean reversion level η	0.02298	0.01346	0.01250
	(0.02068)	(0.01171)	(0.01117)
Speed of mean reversion κ	0.75407	1.41005	1.83351
	(0.68577)	(1.34831)	(1.70350)
Volatility σ	0.12144	0.08924	0.08446
	(0.12069)	(0.08881)	(0.08528)
Jump intensity μ	–	1.46516	0.00000
	–	(1.41013)	(0.00000)
Jump size mean ζ	–	0.00921	0.00760
	–	(0.00917)	(0.00776)
Jump intensity ξ^1	–	–	127.87000
	–	–	(117.18500)
Log-Likelihood	954.46	975.72	980.52
	(963.35)	(984.60)	(989.15)

Table 3.13: Parameter estimates of the portfolio models based on our first estimation approach. Estimates have been calculated by taking the heterogeneity of the portfolio objects into account. Estimates that are obtained under the assumption of homogeneous default intensities are given in parentheses.

risk characteristics of structured credit products.

We have also estimated versions of the CIR, BAJD, and SAJD model under the additional inclusion of a regime process as discussed in Section 3.2. Namely, we specified the mean reversion level η and the jump intensity parameter μ of the factor to be governed by such a process. In this case, model inference was conducted using the Monte-Carlo EM algorithm introduced in the previous subsection. We found that the maximum likelihood was attained in each of the models with a trivial regime process that never changed its state and for the parameter values of Table 3.13.

In the second estimation approach presented in the previous subsection, we cannot directly estimate the portfolio model based on the observed portfolio intensity path and we have to rely on the Monte-Carlo EM algorithm right from the beginning. We estimated three models in which factor and idiosyncratic component evolved according to a CIR, a BAJD, and a SAJD model as suggested by Section 3.2. In

each case, however, we found that the maximum likelihood was attained at $\varpi = 1$ and the parameters reported in Table 3.13 for the assumption of homogeneous default intensities. Note that in our second estimation approach we presume homogeneity of the default intensities, too. A parameter value of $\varpi = 1$ means that the idiosyncratic component $\tilde{\lambda}^{idio} = 0$ and implies that the default risk of each firm stemming from its intensity is purely systemic.

This is surprising but in line with results of Feldhütter (2008). In his empirical investigation of the Duffie and Gârleanu (2001) model, which in our terminology is called the BAJD model, he reports an estimated ϖ parameter of 0.9742. Although he assesses the ability of the model to explain the prices of synthetic structured credit products referencing the CDX index over time, the ϖ value that he obtains does not reflect risk premiums and has to be considered as a real-world measure parameter.

3.5.3 Simulating Transition Matrices of Structured Credit Products

Having estimated the different models, we finally set the stage to answer the question which is at the heart of this section: What are the implications of the model choice for the risk characteristics of structured credit products?

An important risk characteristic of a structured finance product is its rating, because many market participants base their investment decisions on it, and ratings play an important role in the Basel II capital requirements. The rating of a security is usually based on the first moment of its loss distribution. Moody's, for instance, claims consistency of its structured finance ratings with the expected loss (see e.g. Yoshizawa (2003)). A corporate bond and a structured finance security can therefore be identically rated, although their risk characteristics differ. Due to the tranching of claims, the probability of suffering large losses will usually be higher for a structured finance security than for a corporate bond with the same rating. For a typical loss distribution of a tranche see Figure 4.9 on p. 188. Also, the structured finance security will usually be much more exposed to systemic risk than the corporate bond. In the wake of the current crisis, this has been criticized a lot and some market participants now doubt the appropriateness of ratings to describe the risk of structured finance products sufficiently well. It is, however, important to note that this issue has been pointed out earlier, see e.g. Fender and Mitchell (2005).

Methodology

Despite the described shortcomings, the rating is very likely to stay one of the most important publicly available risk measures for structured finance products in the future. We base our investigation on the rating without discussing its appropriateness. Rather, we intend to characterize its dynamics. We proceed as follows:

1. First, we simulate 2000 portfolio loss paths using the estimated models from the previous subsection over a five-year period, which relates to the typical five-year maturity of a CDO. To avoid giving too much weight to the starting value of the factor λ^c, each path uses the end value of the preceding path as starting value.

2. In a second step, we calculate for each path ratings of tranches $[K_l, K_u]$ referencing the simulated portfolio; ratings are calculated year-wise and the attachment and detachment points that we consider are the same six as those of Section 3.4.1: $[0, 0.03]$, $[0.03, 0.06]$, $[0.06, 0.09]$, $[0.09, 0.12]$, $[0.12, 0.22]$, and $[0.22, 1]$. Consistent with Yoshizawa (2003), a tranche's rating at time t is obtained by the following procedure: Given the observed losses in the portfolio up to time t and the realization of the factor we calculate the tranche's expected loss at maturity. The expected loss is then "translated" into a rating using "Moody's Idealized Expected Loss Table" for CDOs as stated in Yoshizawa (2003). In total, this yields $6 \times 2000 = 12,000$ paths of rating history, where each path comprises five, yearly observations. To compute the tranches' expected loss, we apply an approach that is based on Fourier inversion and is described in Appendix A; it guarantees a fast and accurate computation of the expected loss.

3. Based on the generated rating transition data, we finally compute one-year transition matrices.

As previously indicated, our primary focus will be on differences between the simulated transition matrices with respect to the underlying models and between simulated matrices and real-world matrices which have been derived from CDO rating changes in the past.

Results

For illustration purposes, Table 3.14 shows the one-year rating transition matrix which is obtained from our basic regression intensity data set of Section 3.1. It should be mentioned that rating withdrawals were not taken into account in the

computation of the matrix.[28] The obtained matrix resembles the one reported by Moody's Investors Service (2007) for the time period 1984-2006. Our downgrade and upgrade rates presented in Table 3.17 on p. 146 are similar, too.

	Aaa	Aa	A	Baa	Ba	B	Caa	Ca or C
Aaa	0.8686	0.1222	0.0092					
Aa	0.0034	0.8946	0.0972	0.0042	0.0006			
A	0.0002	0.0149	0.9117	0.0653	0.0063	0.0013	0.0002	
Baa		0.0005	0.0402	0.8959	0.0538	0.0083	0.0011	0.0002
Ba			0.0018	0.0488	0.8449	0.1000	0.0044	0.0002
B			0.0008	0.0025	0.0654	0.8744	0.0477	0.0091
Caa					0.0138	0.1077	0.8227	0.0558
Ca or below				0.0062	0.0083	0.0971	0.2397	0.6488

Table 3.14: One-year rating transition matrix of corporates based on our basic regression intensity data set of Section 3.1. Rating withdrawals are not taken into account.

In our study of structured finance rating transitions, we examine two different stylized portfolios: Each portfolio consists of 125 homogeneous firms, but portfolios differ in average creditworthiness: In portfolio I, firms have a low five-year default probability of about 1% and in portfolio II the default probability is about 5%. Different default probabilities are obtained by choosing different a_i values for the intensities (see e.g. equation (3.27)). The factor parametrizations used are those which have been derived in the previous section.

In Table 3.15, we present the one-year CDO rating transition matrix implied by the estimated CIR model and portfolio II showing a high average default rate. The rating transition matrix is stated with respect to the broad rating categories {"Aaa", "Aa", ..., "Caa or below"}. In Table 3.17 on p. 146, the corresponding global rating transition statistics are displayed. Measures such as rating volatility are widely used in order to describe the rating transition behavior. For their definition, we refer to Moody's Investors Service (2007). We find that, excluding the top-level rating category, the rating stability is much lower than the one of the corporates. Although the upgrade rate is very high, downgrade rates in lower rating categories are substantially higher than in the corporate finance transition matrix of Table 3.14.
The observed high upgrade rate has to be attributed to the fact that our estimated models imply an upward-sloping rating "term-structure". This means, for

[28] More precisely, a changed or unchanged rating only enters into the computation of the matrix if the firm has not defaulted during the considered one-year period and all explanatory variables of the firm are available at least at the beginning and the end of this period.

	Aaa	Aa	A	Baa	Ba	B	Caa or below	Default
Aaa	0.9786	0.0213	0.0001					
Aa	0.4692	0.4524	0.0660	0.0107	0.0016			
A	0.0746	0.4970	0.2937	0.1087	0.0229	0.0029	0.0003	
Baa	0.0141	0.1421	0.3149	0.3548	0.1452	0.0232	0.0055	0.0003
Ba	0.0006	0.0140	0.0835	0.3725	0.3577	0.1255	0.0457	0.0006
B		0.0004	0.0078	0.0776	0.3306	0.3250	0.2483	0.0104
Caa or below			0.0002	0.0014	0.0258	0.1363	0.5222	0.3141
Default								1.0000

Table 3.15: Simulated, one-year CDO rating transition matrix based on the estimated CIR model and a reference portfolio with a high default rate of about 5%.

instance, that tranches are rated "Caa2" with respect to their five-year expected loss but would be rated "Baa1" if their one-year expected loss was chosen as rating criterion. For a typical rating term-structure, see Table 3.16. On average, ratings will therefore be upgraded as time evolves leading to higher upgrade rates than downgrade rates.

	Maturity									
Tranche	1	2	3	4	5	6	7	8	9	10
[0.00,0.03]	Caa3	Ca	Ca	Ca	Ca	Ca	Ca	Ca	Ca	Ca
[0.03,0.06]	Baa1	Ba1	B1	B3	Caa2	Caa3	Ca	Ca	Ca	Ca
[0.06,0.09]	Aaa	Aa3	A3	Baa3	Ba1	Ba2	B1	B3	Caa1	Caa2
[0.09,0.12]	Aaa	Aaa	Aa1	Aa2	A1	A3	Baa2	Baa3	Ba1	Ba2
[0.12,0.22]	Aaa	Aaa	Aaa	Aaa	Aa1	Aa1	Aa1	Aa2	Aa3	A1
[0.22,1.00]	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa	Aaa

Table 3.16: Typical rating term-structure presented for the CIR model and portfolio II showing a high five-year average default probability of about 5%. As starting value for the factor process the estimated mean reversion level η was chosen. Here, rating category "Ca" is short for "Ca or below".

Table 3.18 on p. 146 shows the one-year transition matrix based on the SAJD model and the high default portfolio II. A juxtaposition of the matrices and the corresponding global rating transition statistics implied by the different models reveals that differences between the models are small. In the CIR model, rating transitions between the broad rating categories seem to occur even more often than in the SAJD model. Overall, the rating volatility is comparable in both models, but the SAJD model yields a higher upgrade rate.

	Corporates	Structured Finance						
		Low default portfolio I			High default portfolio II			
		CIR	BAJD	SAJD	CIR	BAJD	SAJD	CIR, No Term-Structure
Downgrade Rate	0.1458	0.0653	0.0594	0.0667	0.1746	0.1646	0.1695	0.2555
Upgrade Rate	0.0820	0.1278	0.1171	0.1425	0.3059	0.2439	0.3548	0.0580
Downgrade/Upgrade Ratio	1.7779	0.5113	0.5072	0.4680	0.5706	0.6749	0.4776	4.4017
Downgrade Rate (notch weighted)	0.2637	0.1619	0.1529	0.1601	0.3181	0.2755	0.2840	0.8244
Upgrade Rate (notch weighted)	0.1148	0.1762	0.1315	0.2017	0.7097	0.4539	0.7418	0.1097
Downgrade/Upgrade Ratio (notch weighted)	2.2981	0.9188	1.1627	0.7936	0.4482	0.6071	0.3828	7.5141
Rating Drift (notch weighted)	-0.1490	0.0143	-0.0214	0.0416	0.3916	0.1784	0.4578	-0.7146
Rating Volatility (notch weighted)	0.3785	0.3382	0.2844	0.3618	1.0278	0.7294	1.0257	0.9341
Rating Stability	0.7721	0.8069	0.8235	0.7909	0.5195	0.5914	0.4757	0.6865

Table 3.17: Global rating transition statistics. Statistics are stated for a low default portfolio with a five-year default probability of about 1% and a high default portfolio with a default probability of 5%. Rating transitions are based on 2000 simulated paths of our three estimated portfolio models (CIR, BAJD, SAJD). "No Term-Structure" means that the rating is in this case determined based on the one-year ahead expected loss instead of the expected loss at maturity.

	Aaa	Aa	A	Baa	Ba	B	Caa or below	Default
Aaa	0.9924	0.0076						
Aa	0.3836	0.5800	0.0326	0.0035	0.0003			
A	0.0533	0.4209	0.4050	0.1070	0.0123	0.0013	0.0003	
Baa	0.0032	0.0931	0.4459	0.3440	0.0980	0.0135	0.0022	
Ba		0.0036	0.0581	0.3283	0.4599	0.1197	0.0297	0.0008
B			0.0035	0.0614	0.4077	0.3203	0.1980	0.0091
Caa or below				0.0003	0.0162	0.1091	0.5807	0.2937
Default								1.0000

Table 3.18: Simulated, one-year CDO rating transition matrix based on the estimated SAJD model and a reference portfolio with a high default rate of about 5%.

	Aaa	Aa	A	Baa	Ba	B	Caa or below	Default
Aaa	0.9199	0.0580	0.0163	0.0049	0.0008	0.0001		
Aa	0.1846	0.3187	0.2697	0.1632	0.0451	0.0163	0.0023	
A	0.0126	0.1124	0.2963	0.3349	0.1688	0.0548	0.0193	0.0008
Baa	0.0007	0.0180	0.1330	0.3175	0.3085	0.1495	0.0692	0.0036
Ba		0.0022	0.0139	0.1314	0.3156	0.2596	0.2357	0.0416
B				0.0190	0.1188	0.2834	0.4321	0.1467
Caa or below					0.0013	0.0201	0.3388	0.6398
Default								1.0000

Table 3.19: Simulated, one-year CDO rating transition matrix based on the estimated CIR model, a reference portfolio with a high default rate (five-year default probability of about 5%) and ratings determined according to the expected loss at the one-year horizon.

	Maturity									
Tranche	1	2	3	4	5	6	7	8	9	10
[0.00,0.03]	Caa2	Caa2	Caa2	Caa2	Caa2	Caa3	Caa3	Caa3	Caa3	Caa3
[0.03,0.06]	B2	B2	B2	B2	B2	B2	B2	B3	B3	B3
[0.06,0.09]	B1	B1	Ba3	Ba3	Ba3	Ba3	B1	B1	B1	B1
[0.09,0.12]	Ba2	Ba3	Ba3	Ba2	Ba2	Ba2	Ba3	Ba3	Ba3	Ba3
[0.12,0.22]	Baa3	Ba1	Ba1	Ba1	Ba1	Ba1	Ba1	Ba1	Ba1	Ba1
[0.22,1.00]	Aa3	A3	A3	A2	A2	A2	A2	A2	A2	A2

Table 3.20: Rating term-structure in case of the adjusted CIR model ($\eta = 0.00715$, $\kappa = 0.754$, $\sigma = 0.714$) a portfolio with a default probability of about 3.1%. As starting value for the factor process the estimated mean reversion level η was chosen. Model parameters were determined such that the rating of each security is the same for as many maturities as possible.

	Aaa	Aa	A	Baa	Ba	B	Caa or below	Default
Aaa								
Aa		0.0483	0.0322	0.9115	0.0054	0.0027		
A		0.0349	0.0484	0.8817	0.0188	0.0108	0.0054	
Baa		0.0145	0.0156	0.9219	0.0268	0.0112	0.0096	0.0004
Ba			0.0003	0.0313	0.9026	0.0303	0.0296	0.0059
B			0.0003	0.0244	0.0678	0.7765	0.0386	0.0924
Caa or below				0.0427	0.3205	0.2054	0.0345	0.3969
Default								1.0000

Table 3.21: Simulated, one-year CDO rating transition matrix based on a CIR model which implies an almost flat rating term-structure and ratings determined according to the expected loss at the one-year horizon.

Compared with the influence of the model choice, the effect of considering a different underlying portfolio is much stronger. When considering a low default portfolio, the rating stability strongly increases and downgrade and upgrade rates generally

go down. The decrease of the downgrade rate can be explained by the diminished number of defaults which is observed in the low default portfolio, because at a default the expected losses of all tranches jump up often resulting in downgrades. More defaults therefore mean higher downgrade rates. The decreased upgrade rate might be puzzling at first, but note that in the low default portfolio many more securities are rated "Aaa".

As previously pointed out, there is a mismatch between the rating term-structure implied by our estimated models and the model which is at the bottom of Moody's ratings. In order to further investigate this issue, we first remove the term-structure effect from our ratings by determining the rating based on the expected loss after a fixed future time period instead of the expected loss at maturity. We choose the one-year horizon because of the five-year maturity of the securities.

Tables 3.17 and 3.19 present the global rating statistics and the corresponding rating transition matrix based on the estimated CIR model and the high default portfolio II. We find that the term-structure mismatch has a strong effect on the matrices: As soon as it is removed, the rating stability increases, the upgrade rate strongly goes down and the downgrade rate increases. Also, the rating drift becomes negative meaning that on average ratings are downgraded now, and the probability of a default in a low rating category is higher. The Downgrade/Upgrade ratio is larger than the one of corporate finance, but still below the 7.47 reported by Moody's Investors Service (2007) for CDOs over the time period 1997-2006.

Another possibility of eliminating the term-structure effect is to choose the parameters of our model such that the ratings of the considered securities are equal for all maturities: We keep the speed of mean reversion fixed at $\kappa = 0.754$ and adjust the mean reversion level η and the volatility σ of the CIR model, which yields $\eta = 0.00715$ and $\sigma = 0.714$. The corresponding rating term-structure is displayed in Table 3.20. Although the derived parameters imply a much stronger dependence structure between the default events than our parameters estimated based on the portfolio intensity, the implied term-structure is not completely flat. This can be attributed to the fact that dependencies between the portfolio objects decrease for shorter maturities, because the considered models exclude simultaneous defaults. Table 3.21 shows the transition matrix derived with the adjusted CIR model. On the one hand, we find that there are high downgrade probabilities in the high and in the low rating categories. On the other hand, "Baa", "Ba" and "B" ratings are relatively stable.

In total, these results indicate that we have to be careful with definite conclusions because of the problems that we encountered during our investigation. The problems were primarily related to the fact that our estimated models do not imply

consistent ratings for all considered maturities, and the missmatch has a strong impact on the results. It is possible to remove this effect by computing a security's rating based on the one-year ahead expected loss and not based on the expected loss at maturity, for example, but it is not clear whether this is appropriate. In particular, it contradicts the rating methodology of Moody's (see Yoshizawa (2003)).

A common feature of all structured finance rating transition matrices which have been derived in our investigation is that they implied a rating stability that decreases in the riskiness of the underlying portfolio. High default rates in this portfolio imply a higher overall rating volatility and higher downgrade rates, a fact, which should be taken into account by investors. Also, the simulated rating transition data suggests a high default probability of structured finance products having a low rating. In comparison with these general effects the consequence of changing from one model to another was only small, which reduces the possibility of high model risk regarding the considered models.

3.6 Summary and Remarks

In this chapter, we investigate the question of which type of models are needed in order to model the dynamics of credit portfolios that are observed in real data. The implications of our findings for structured credit modeling are discussed. We start by estimating default intensities for a large sample of US and non-US corporates. In contrast to Das et al. (2007), we show that estimated default intensities are able to explain the observed default clustering, although we estimate the intensities based on observable covariates such as the firms' Expected Default Frequency (EDF) and do not introduce additional contagion effects or "frailty" variables. We modify their estimation approach in two ways. First, we model intra-month patterns in observed defaults. Second, we estimate default intensities on an out-of-sample basis, which brings our estimates closer to the ones financial institutions implementing the models would actually have used. Once intensity estimation is modified in these ways, the hypothesis of well-specified intensities is no longer rejected. In addition, when examining the ability of our regression intensity model to rank firms according to their default likeliness we find that its predictive power is higher than that reported by Duffie et al. (2007) for their regression intensity model and a similar data set.

Subsequently, we introduce a time-continuous model in order to explain the joint dynamics of the intensities. The model includes other established models in literature as special cases. In the simplest case, intensities follow a *Cox-Ingersoll-Ross* process (see Cox et al. (1985)). In general, the model represents a solid basis for our empirical investigation since it exhibits a high degree of analytical tractability,

and the different nested models can easily be compared.

In the first application of the model, we compare the ability of the different model versions to explain the intensity of each single firm and examine the predictive power of the models. Especially for firms of bad creditworthiness, we find that models with intensity jumps are better able to model the intensity dynamics than purely diffusion-based models. Concerning default prediction, however, we find that more complex models do not yield better results.

Afterwards, we investigate the issue of model risk in connection with the assumption of conditional independence. Our simulation study reveals that estimation errors are by far more influential than errors related to this assumption. Even though contagion effects played a dominant role in the data generating model, some of the conditional independence models "lead" to results that are similar to those obtained from the estimated true model.

Towards the end of this chapter, we present the model estimation on a portfolio basis. We compare the estimated model versions with respect to their ability to explain the portfolio intensity. In a second step, using the estimated parameters we simulate paths of the portfolio loss process and the corresponding portfolio intensity. In this way, we can compute a time series of ratings for hypothetical, structured credit products referencing the portfolio and eventually obtain transition matrices for these products. This allows us to compare the different model versions with respect to these matrices. We find that simple and complex models eventually imply comparable risk profiles of structured credit products.

In summary, the findings of this chapter show that – as expected – simple and more complex models yield different results. However, the influence of different specifications on the results is minor in our data set when compared with general problems that apply to all model versions. In particular, estimation errors are considerable and their size usually increases in the number of parameters to be estimated. For practical applications such as the risk analysis of structured credit products, the issue of how to reduce these estimation errors is therefore by far more important than the model choice and needs to be solved first – if even possible – before considering more complex models.

Chapter 4

Application II: Analysis of Mortality Contingent Catastrophe Bonds

As mentioned earlier in this thesis, (re)insurance companies are exposed to different sources of risk: Changes in interest rates, people who live systematically longer than expected or catastrophe events like earthquakes or hurricanes could cause large losses. To actively manage these risks on their balance sheets, (re) insurance companies have started to transfer some of them to the capital markets – similar to the way how banks deal with their credit risk exposure. By means of securitization, i.e. isolating the cash flows that are linked to insurance liabilities and repackaging them into cash flows that are traded in capital markets, insurers and reinsurers have begun to step away from their traditional *risk warehousing* function towards a business model of *risk intermediation*.

One prominent example of *Insurance Linked Securities* (ILS) which enables (re) insurers to load extreme tail risks off their balance sheets are so-called *Catastrophe* (CAT) Bonds, the coupons and principal payments of which depend on the incurrence of certain catastrophic events. They have been traded since the mid 1990s and present interesting investment possibilities as they are "low-beta" investments, i.e. their returns show a low correlation to financial markets, and thus increase diversification possibilities for investors (see Cox et al. (2000)). Moreover, they offer several potential advantages over alternative methods for insurers to deal with catastrophic risk. For instance, in comparison with traditional (re) insurance catastrophe bonds bear less credit risk (see Niehaus (2002)).

The market for CAT securities and the pricing of CAT bonds have been studied in various contributions (see e.g. Doherty (1997), Froot (2001) and Lee and Yu (2002), Young (2004), respectively). However, the amount of capital that has been

raised within securitizations of catastrophic risk remains small (Cummins (2006)). One possible explanation is that these bonds may be expensive relative to conventional reinsurance since investors often charge a high risk premium on the bonds (cf. Bantwal and Kunreuther (1999), Froot (2001)).

A more recent capital market innovation are so-called *CAT Mortality Bonds* (henceforth CATM bonds). While most CAT bonds and other CAT derivatives depend on underlying loss indices such as the Property Claim Services (PCS) loss index, CATM bonds are contingent on less artificial events: They are triggered by a catastrophic evolution of death rates of a certain population. Investors' demand for these securities seems to be very high, and in contrast to "conventional" CAT bonds the number of deals has increased considerably over the last years: There were four major deals in 2006 with a total volume of more than \$1.2 Billion. Surprisingly, there have been only very few contributions in the scientific literature on CATM bonds.

Lin and Cox (2006) and Cox et al. (2006) develop an asset pricing model for mortality contingent securities in an incomplete market framework with jump processes. In particular, they propose a pricing method for CATM bonds. Modeling the underlying combined mortality index which triggers the bond by a geometric Brownian motion with a multiplicative jump component and distorting the resulting distributions by the so-called *Wang transform* (see Wang (2000) and Wang (2002)), they are able to explain market outcomes of existing mortality securitizations regarding investors' demand by analyzing the implied risk premiums. However, their pricing model remains static in the sense that no risk-adjusted dynamics are derived. Furthermore, they only focus on a single transaction, namely the first one, and do not provide an overview on the CATM market thus far; particularly, the stability of their findings regarding subsequent transactions is not examined. Also, modeling the index rather than the evolution of the underlying mortality may not be adequate as the index corresponds to one specific deal.

The present chapter aims to close this gap in the literature. Aside from providing a concise overview on the market history, we develop a risk assessment and pricing model which is based on stochastic modeling of the mortality intensity (cf. Subsection 2.8.2). Parametrizations of the proposed model based on three different calibration procedures are derived; we provide the resulting loss profiles as well as prices, and discuss the consequences of our findings.

The remainder of this chapter is organized as follows. After explaining the general structure of CATM bonds based on the so-called Tartan bond[1] in Subsection 4.1.1,

[1] *Tartan Capital Ltd. Series 1* arranged by *Goldman Sachs* for the reinsurer *Scottish Re Group Ltd.* (Scottish Re), issued in May 2006.

Subsection 4.1.2 provides an overview on the development of the CATM securitization market thus far. In particular, we present the characteristics of all deals until the end of 2006. After a short overview in Subsection 4.1.3 on general modeling approaches with a focus on those used in practice, we introduce in 4.2 our model based on the stochastic mortality modeling approach described in Subsection 2.8.2. Our model consists of two components: A baseline component governed by a diffusion reflecting the "regular" fluctuations of mortality over time and a catastrophe component driven by a jump process representing catastrophic events. Three calibration procedures for this model are presented in Section 4.3: First, best estimate parametrizations based on historical data and viewpoints from the demographic literature are derived; the other two risk-adjusted parametrizations are extracted from market prices of term life insurance policies and past catastrophe mortality securitizations, respectively. Section 4.4 presents our results; we provide prices in terms of excess spread levels, loss probabilities as well as expected losses corresponding to the parametrizations derived in Section 4.3 and compare our results to loss profiles provided by the issuers and market prices. Moreover, similarities to the credit risky securities considered in Chapter 3 are pointed out. After a discussion of our findings, Section 4.5 concludes.

4.1 An Overview on Catastrophe Mortality Bonds

Securitization transactions are usually highly complex and involve several parties such as lawyers, rating agencies, trustees, etc. Providing a general overview on securitization transactions is far beyond the scope of this thesis (see e.g. Deacon (2006) or Jeffrey (2006) for an introduction to securitization and *Asset Backed Securities* (ABS) in general, or Cowley and Cummins (2005) for life insurance securitization). The basic idea is to isolate and pool cash flows that are linked to certain assets or liabilities, repackage them into cash flows which support certain related securities and issue these securities to capital markets.

Within CATM securitizations, insurers and reinsurers transfer catastrophe mortality risk, which arises from a possible occurrence of, for example, severe pandemics or catastrophic terrorist attacks from their liability side to the capital market by means of CATM bonds. Traditionally, these risks were shared between insurers and reinsurers via reinsurance or retrocession. However, in contrast to these classical approaches, securitization avoids credit risk (see also Niehaus (2002)). Moreover, a traditional risk transfer may be more expensive as possible transaction partners usually already have this type of risk in their books and thus their appetite for it is limited. Also, retrocession would require the disclosure of the own business to possible competitors.

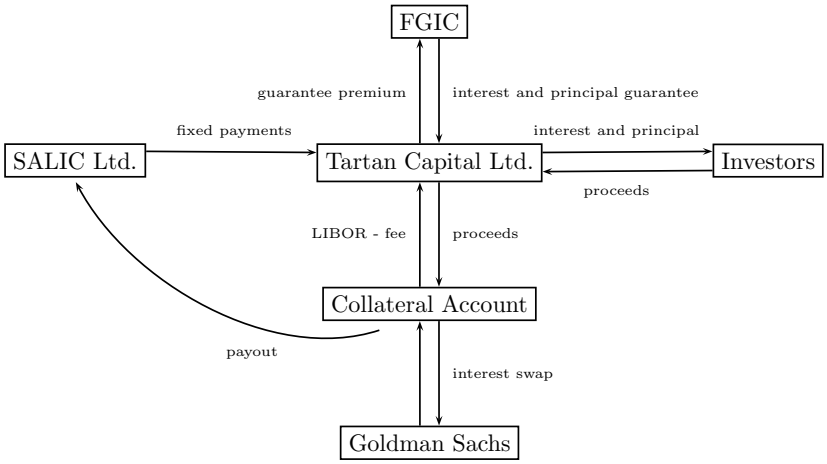


Figure 4.1: Description of the Tartan deal structure (Source: Linfoot (2007)).

Thus far, there have been five public transactions. While they differed in their coverage area, credit ratings, or spread levels, the basic structure is the same: A certain underlying mortality index based on the mortality experience in the coverage area is defined; if this index exceeds certain pre-specified levels, the bond is triggered, i.e. the investors start to loose their principal. As it is cumbersome to present the characteristics of all available CATM bonds at the same time, in Subsection 4.1.1, we detail out the structure of one representative example, namely the third of all five transactions: the Tartan transaction arranged by Goldman Sachs for the reinsurer Scottish Re (cf. Linfoot (2007)). In Subsection 4.1.2, we then compare this deal to the other transactions so far.

4.1.1 Structure of the Securities

In Figure 4.1, the structure of the Tartan transaction is illustrated. SALIC[2], a member of the *Scottish Re Group Ltd.*, entered into a counter-party agreement with the special purpose vehicle *Tartan Capital Ltd.* (Tartan). Under this agreement Tartan is obligated to make payments to SALIC in case a certain index is triggered. In return, SALIC agreed to pay Tartan a certain fixed amount quarterly. In order to raise funds for the conditional payments to SALIC, Tartan issued and sold bonds to capital market investors; the proceeds were used to buy eligible securities which act as collateral. As these collateral assets could decrease in market value, Tartan went into a swap agreement with Goldman Sachs, who have also structured the deal: In return for the variable investment income from the collateral account, Goldman Sachs agreed to pay the 3-month LIBOR[3] minus a fee of 10 basis points (bps).

[2] *Scottish Annuity & Life Insurance Company (Cayman) Ltd.*
[3] *London Interbank Offered Rate.*

Tartan issued 2 series of 3-year notes: a \$75 million (mn) (Class A) and an \$80 mn (Class B) tranche with different risk exposures. In particular, within the Class A notes both interest payments and the investors' principal are guaranteed by the monoline insurer *Financial Guaranty Insurance Co.* (FGIC). Therefore, the only risk that investors in the Class A notes have to face is credit risk. In return for the guarantee, FGIC received a premium from Tartan. Class B investors, on the other hand, are actually exposed to catastrophe mortality risk, i.e. they will lose interest and principal in case of a trigger event.

The bonds and thus the payment to SALIC are triggered if a well defined parametric index exceeds a certain level. This so-called *combined mortality index* (CMI) is contingent on the mortality experience of certain populations, and the objective is to design it such that the actual catastrophe mortality exposure of the protection buyer (SALIC / Scottish Re) is reflected as well as is possible. Within the Tartan transaction, this index is solely based on US population mortality. For each relevant point in time (calendar year) t, the mortality rates, i.e. the probabilities to decease within the following year for certain partitions of the whole population as reported from the *Centers for Disease Control and Prevention* (CDC), are weighted to determine a *weighted population death rate* $\hat{q}_t$:

$$\hat{q}_t = \sum_{\text{all } x} \omega_{x,m}\, \hat{q}_{m,x,t} \;+\; \omega_{x,f}\, \hat{q}_{f,x,t}, \tag{4.1}$$

where $\hat{q}_{m,x,t}$ and $\hat{q}_{f,x,t}$ are the mortality rates for age group x in calendar year t for males and females, respectively, and $\omega_{x,m}$ / $\omega_{x,f}$ are the weights applied to the corresponding mortality rates. The weights for the Tartan transaction are displayed in Table 4.1. Now, the actual index at time t, say i_t, is derived from the underlying weighted population death rates at times t and $t-1$ as well as the weighted population death rates for the *reference years* 2004 and 2005, which are determined according to equation (4.1), by the relationship

$$i_t = \frac{\frac{1}{2}\left(\hat{q}_t + \hat{q}_{t-1}\right)}{\frac{1}{2}\left(\hat{q}_{2005} + \hat{q}_{2004}\right)}. \tag{4.2}$$

Since the index relies on the experience of two consecutive years and since Tartan issued bonds with a three year tenor, there are only two dates at which the index is calculated and at which the principal may be reduced due to a potential catastrophic event: at the end of 2007 for the years 2006 and 2007 and at the end of 2008 for the years 2007 and 2008. In particular, this implies that investors cannot lose principal in the first two years. However, the data for the index calculation will usually not be available until a while after the respective measurement dates. Therefore, Tartan has the possibility to extend the tenor of the notes up to a maximum of 30 months, but the securities cannot suffer any losses due to a possible

Age Groups (x)	Age Weights: Male ($\omega_{x,m}$)	Age Weights: Female ($\omega_{x,f}$)
1–4	0%	0%
5–14	0.1%	0.1%
15–24	0.4%	0.4%
25–34	8.2%	6.1%
35–44	26.0%	12.7%
45–54	21.4%	7.8%
55–64	9.8%	2.7%
65–74	2.3%	0.8%
75–84	0.6%	0.4%
84+	0.1%	0.1%
Total	68.8%	31.2%

Table 4.1: Gender and age weights for the Tartan transaction (Source: Linfoot (2007)).

event within the extension period, and investors will receive ongoing interest payments.

Furthermore, only if the index exceeds a certain level, the so-called *trigger level* or *attachment point* a, investors will lose principal. If the index exceeds the so-called *exhaustion level* or *detachment point* d, their complete principal will be lost. For index levels between the attachment and detachment points, the loss percentage of the principal at time $t = 2007, 2008$, l_t, is determined as follows:

$$l_t = \min\left\{\max\left\{l_{t-1}, \frac{i_t - a}{d - a}\right\}, 100\%\right\}, \tag{4.3}$$

where $l_{2006} := 0$. Coupons are only paid on the remaining principal. Table 4.2 provides the trigger and exhaustion levels as well as the interest on the Tartan notes.

Interest on the notes is paid quarterly. It is worth mentioning that the spread levels are not fixed from the beginning of the marketing phase of the notes – they depend on investors' demand and market conditions.

[4]Rating at Issuance from *Moody's Investors Service* (Moody's) and *Standard and Poor's* (S&P).

	Class A Notes	Class B Notes
Tranche Size	\$75mn	\$80mn
Term	3 years	3 years
Trigger Level	115%	110%
Exhaustion Level	120%	115%
Coupon (bps)	LIBOR+19	LIBOR+300
Rating[4]	Aaa/AAA	Baa3/BBB

Table 4.2: Program summary of the notes issued by Tartan (Source: Linfoot (2007)).

4.1.2 Market Development[5]

As mentioned earlier in this section, Tartan was the third public catastrophe mortality transaction. Table 4.3 provides an overview on all such transactions so far.

Vita Capital Ltd. (Vita I) was the first CATM securitization transaction out of the \$2 billion (bn) multi-currency shelf program[6] established by *Swiss Reinsurance Company* (Swiss Re). While the latter four deals are still ongoing, Vita I matured in the end of December 2006 and was not extended as there was no extreme mortality event during the risk period. In contrast to the Tartan deal, it only had one single tranche and the underlying index was based on the population from several countries rather than just one.[7] When structuring the transaction, Swiss Re wanted to involve an American monoline insurer to "wrap" the bond, but eventually due to regulatory issues no cooperation was established. Swiss Re managed the challenge of selling this new type of risk to the market, and the notes were placed successfully.

About 18 months after Vita I, again Swiss Re came to the market with their second transaction, *Vita Capital II Ltd.* (Vita II). In contrast to Vita I, Vita II has three tranches with different, decreasing seniorities due to decreasing trigger and exhaustion levels, but all of them are of a lower seniority than the single Vita I tranche. Despite this fact, the spread level within the first transaction exceeds the

[5]Based on Logisch (2007).

[6]Shelf program means that not the total capacity is issued initially – some "sits on a shelf". Establishing a shelf program reduces costs for future transactions as the legal work, modeling, etc. are done for a relatively large amount. Moreover, it reduces the time from the decision to access the capital market and the closure of the deal enabling the issuer to quickly react when protection is needed or investors' appetite is large enough to absorb the extra issued bonds.

[7]In this case, the combined mortality index is defined as the weighted average over the indices from the individual countries determined according to equation (4.2) with country weights as provided in Table 4.3.

level of the Class B note from Vita II. This tight pricing was possible since the bond was over-subscribed indicating that within Vita I investors had demanded a considerable novelty premium, i.e. an additional premium for the unknown asset class. However, Swiss Re did not seem to be surprised by this fact since they chose to issue the riskier and potentially more expensive tranches after the market had got acquainted with this type of security. As indicated in Table 4.3, S&P upgraded all three Vita II classes by one notch in April 2006. According to Standard and Poor's (2006), this upgrade was mainly due to the availability of new mortality data showing mortality improvements, advances in vaccine research and continuing work of governments regarding their pandemic preparedness plans.

As mentioned above, Tartan was the third CATM transaction, and so far the only one without involvement of Swiss Re. This first issue out of Scottish Re's $300mn shelf structure was also the first issue with a tranche wrapped by a monoline insurer. In comparison to the Vita II Class D tranche, the non-guaranteed Class B note is priced considerably higher even though both have the same trigger and exhaustion levels. This may be due to the fact that, in contrast to the deals before, Tartan is solely based on US mortality experience and, thus, there is no diversification effect among several populations. However, the more important reason was likely timing: During the marketing period of the transaction in the beginning of 2006, the international press had paid an increased attention to possible outbreaks of Avian Flu and to pandemics in general (see e.g. the *Pandemic Theme Index* provided by *Conquest Investment Advisory AG*).

Six months after Tartan, once the public discussions regarding pandemic fears had calmed down, the fourth series of CATM bonds were issued by *OSIRIS Capital Plc* (OSIRIS). Again, Swiss Re was involved as structurer and lead underwriter for the underlying EUR1.0bn shelf program but not as the protection buyer; the program has been structured as a securitization for the catastrophe mortality risk within the books of *AXA Cessions* (AXA), a subcompany of the French *AXA group*. Therefore, it is the first deal which involves a primary insurer, and for the first time the underlying CMI is not dominated by US mortality experience. Aside from Swiss Re, *IXIS Corporate and Investment Bank* (IXIS) and *Lehman Brothers Inc.* (Lehman) were invited to act as co-underwriters. According to a press release from Swiss Re on 11/13/2006, investors' demand was very strong and all classes were oversubscribed. Euroweek (2006) even reports that all tranches were increased in size due to high investor demand and that all classes were priced well within the price guidance. However, even though the Baa2/BBB rated Class C tranche was priced tighter than the Baa3/BBB+ Class B note from Tartan, its price is still far from the level of the similar Vita II Class D Bond. This may be a consequence of increased investors' expectations after the Tartan transaction. The high demand was mainly due to the fact that in addition to specialized CAT bond investors Swiss

Re, IXIS, and Lehman also approached traditional ABS investors. Furthermore, with the Ba1/BB+ rated Class D note a non-investment grade tranche was offered within a CATM securitization for the first time. Thus, the deal has also drawn attention from hedge funds. All in all, 50% of the bonds were sold to asset managers, 20% to banks, and 25% to hedge funds (cf. International Financing Review (2006)).

In late December of 2006, *Vita Capital III Ltd.* (Vita III) issued the third series of bonds out of Swiss Re's shelf program with the key objective to replace Vita I, which had expired in the same month. However, the total amount was increased in comparison to Vita I. All offered tranches are of a high seniority with all ratings above A, and they were priced similarly to the comparable OSIRIS notes. It is worth noting that five of the nine tranches are wrapped by three different monoline insurers.

In conclusion, it appears that in the relatively short history of the CATM market the spectrum of investors has broadened considerably. While the initial transactions were mainly geared towards specialized CAT bond investors, more and more fixed-income and traditional ABS investors seem to be interested. This may be reasoned with the low correlation or the *one-way* relationship with debt capital and equity markets and the resulting diversification possibilities, but the attractive risk-return profile when comparing CATM bonds to similar rated Mortage Backed Securities (MBS) or Collateralized Debt Obligation (CDO) tranches certainly plays a role, too.[8] For example, an anonymous investor explained that he is investing in CATM bonds because of the relative high spread margins and added: *"If there will be one day such a severe world-wide pandemic that one of the bonds I bought will be triggered, there will be more important things to look after than an investment portfolio."*

In order to estimate and analyze the risks within CATM bonds, investment managers started hiring actuaries to act as specialists on insurance risks. However, so far the market participants mainly rely on the advice of so-called modeling firms. In the next section, after providing an overview of these consultants and their modeling approaches we introduce a model which can be used to price and analyze extreme mortality risks.

[8]The term *one-way* relationship means that adverse events in the financial market have no impact on the performance of a CATM bond, whereas a severe pandemic could affect the financial markets considerably.

	Vita Capital Ltd.	Vita Capital II Ltd.			Tartan Capital Ltd.	
Issued	Nov. 2003	Apr. 2006			May 2006	
Class[9]	A	B	C	D	A*	B
Tranche Size	$400mn	$62mn	$200mn	$100mn	$75mn	$80mn
Arranger	Swiss Re		Swiss Re		Goldman Sachs	
Protection for	Swiss Re		Swiss Re		Scottish Re	
Rating[10]	A3/A+	Aa3/A-**	A2/BBB+**	Baa2/BBB-**	Aaa/AAA	Baa3/BBB+
Attachment Point	130%	120%	115%	110%	115%	110%
Detachment Point	150%	125%	120%	115%	120%	115%
Coupon (bps)	LIBOR+135	LIBOR+90	LIBOR+140	LIBOR+190	LIBOR+19	LIBOR+300
Expected Maturity	4 years	5 years	5 years	5 years	3 years	3 years
Covered Area	US 70%, UK 15%, F 7.5%, I 5%, CH 2.5%	US 62.5%, UK 17.5%, D 7.5%, J 7.5%, CAN 5%			US 100%	

	Osiris Capital Plc.			
Issued	Nov. 2006			
Class	B1*	B2	C	D
Tranche Size	Euro 100mn	Euro 50mn	$150mn	$100mn
Arranger	Swiss Re			
Protection for	AXA			
Rating	Aaa/AAA	A3/A-	Baa2/BBB	Ba1/BB+
Attachment Point	114%	114%	110%	106%
Detachment Point	119%	119%	114%	110%
Coupon (bps)	EURIBOR+20	EURIBOR+120	LIBOR+285	LIBOR+500
Expected Maturity	4 years	4 years	4 years	4 years
Covered Area	F 60%, J 25%, US 15%			

	Vita Capital III Ltd.								
Issued	Dec. 2006								
Class	A-IV*	A-V*	A-VI*	A-VII	B-I	B-II	B-III	BV*	BVI*
Tranche Size	$100mn	$100mn	Euro 55mn	Euro 100mn	$90mn	$50mn	Euro 30mn	$ 50mn	Euro 55mn
Arranger					Swiss Re				
Protection for					Swiss Re				
Rating	Aaa/AAA	Aaa/AAA	Aaa/AAA	Aa2/AA-	A1/A	A1/A	A1/A	Aaa/AAA	Aaa/AAA
Attachment Point	125%	125%	125%	125%	120%	120%	120%	120%	120%
Detachment Point	145%	145%	145%	145%	125%	125%	125%	125%	125%
Coupon (bps)	LIBOR+21	LIBOR+20	EURIBOR+21	EURIBOR+80	LIBOR+110	LIBOR+112	EURIBOR+110	LIBOR+21	EURIBOR+22
Expected Maturity	4 years	5 years	4 years	5 years	4 years	5 years	4 years	5 years	4 years
Covered Area	US 62.5%, UK 17.5%, D 7.5%, J 7.5%, CAN 5%								

Table 4.3: Comparison of all CATM deals from 2003 until 2006 (Source: New Issue Reports from S&P and Moody's; *Bloomberg* data).

[9]The tranches marked with * are guaranteed by monoline insurers.

[10]Rating at Issuance from Moody's / S&P – the ratings marked with ** were upgraded by S&P.

4.1.3 Modeling Approaches in Practice

Aside from the arranger, SPV managers, rating agencies etc., so-called *risk modeling firms* play an important role in a catastrophe mortality securitization. They are appointed to calculate loss probabilities and expected losses for the different tranches of a transaction. These loss profiles are important as investors and rating agencies usually base their decisions on this data. Furthermore, they are in charge of calculating the combined mortality index; thus, they are also referred to as *calculation agents*. To date, the global acting actuarial consultant *Milliman Inc.* (Milliman) was hired as the calculation agent in all transactions so far. However, within the Vita III transaction the US based company *Risk Management Solutions* (RMS) was also involved as an adviser for the monoline insurer *Financial Security Assurance Inc.* (FSA). The modeling approaches of Milliman and RMS differ considerably: While Miliman bases its analysis on an actuarial model, RMS uses an epidemiological approach. Although no mathematical details on their respective models can be presented as to our knowledge these are not published, an overview of their approaches based on the available information is provided in Subsection 4.1.3. In Subsection 4.2, we present our approach based on stochastic mortality modeling.

RMS: An Epidemiological Approach[11]

RMS reports that its catastrophe mortality model is based on epidemiological data and research rather than historical data, and that it was supported by world-wide experts in influenza research when developing its methodology. However, historical data is used to test the model. It is worth mentioning that RMS's model does not include man made catastrophes such as terrorism acts.

Within its model, event tree techniques are applied to produce 1,890 probability weighted scenarios; Figure 4.2 shows the underlying event tree. Regarding the probability of an outbreak, RMS mentions that there have historically been an average of three pandemics per century. Thus, RMS assumes an annual outbreak probability of 3-4%. However, as industrialized livestock husbandry and other conditions fostering mutations of viruses have increased in recent decades it is possible that the situation has worsened. Furthermore, the ongoing scientific debate of whether the risk of a pandemic is increased due to H5N1 prevalence in bird populations (the so-called Avian Flu) is another indication that the historical probability may be too low. Therefore, RMS advises using levels of 5% and 6.7% for stress-testing purposes.

[11]Based on Logisch (2007), who reports a web presentation held by an RMS modeling expert in January 2007 as his primary source.

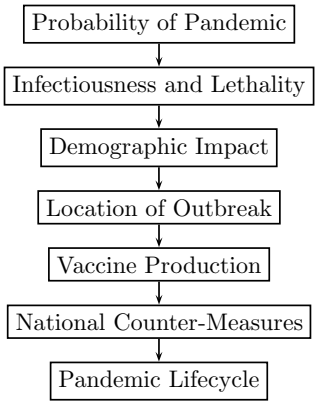


Figure 4.2: RMS Pandemic Influenza Model Framework (Source: Logisch (2007)).

The parameters determining infectiousness and lethality are fixed based on influenza research. In particular, RMS researched what proportion of the population is susceptible to a virus, which proportion will be affected and what the corresponding recovery rates are. More precisely, the demographic impact of a virus is considered in its model as usually the very young and the elderly – those with a weaker immune system – are most affected by a virus. However, due to so-called "cytokine storms", i.e. potentially fatal immune reactions, a strong immune system may rather be a disadvantage than an advantage. For example, within the Spanish Flu from 1918, the most severe pandemic in the last century a disproportionate amount of young adults had been killed, which is believed to be the consequence of cytokine storms. Furthermore, human deaths from H5N1 usually involve cytokine storms.[12]

The location of an outbreak is another important influence factor within the RMS model: Five world regions are being distinguished, and a regional as well as an interregional spread rate is modeled. For example, the underlying data includes maps of the connections between international airports. According to RMS, its spread model yields results which are consistent with the models used by the US government, the UK government and the World Health Organization. The location of the outbreak also determines the effectiveness of the country-specific emergency plans and countermeasures as well as vaccine production scenarios and their probabilities.

RMS claims consistency of its model with all historic data. In its opinion, a severe pandemic would arise from the combination of a virulent and infectious virus, a high incidence of cytokine storms, a delayed vaccine and a failed government

[12]From *www.wikipedia.org*, 02/09/2007.

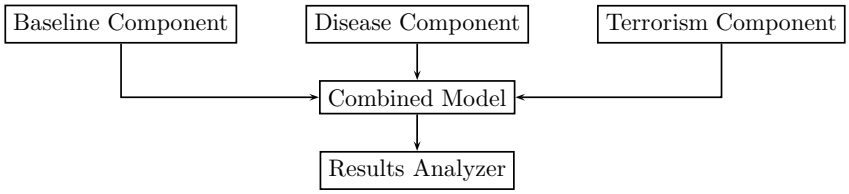


Figure 4.3: Milliman model overview (Source: Logisch (2007)).

response.

Milliman: An Actuarial Approach[13]

As noted above, Milliman's approach differs considerably from the one presented by RMS as Milliman models the future evolution of mortality with actuarial and statistical methods based on historical data. As shown in Figure 4.3, its framework consists of three basic components:

1. The *Baseline Component* models random fluctuations within annual mortality rates as long as no catastrophic event occurs. Using time-series models, Milliman develops stochastic forecasts of the mortality evolution for every country covered in the combined mortality index, which are applied to produce simulations of the weighted combined death rates (cf. equation (4.1)).

2. The *Disease Component* captures the excess mortality due to a pandemic outbreak. The frequency and severity of an outbreak are modeled separately based on data from past pandemics. The same model is used for each relevant country, and it is assumed that pandemics occur simultaneously in these countries. Furthermore, Milliman assumes that these events are independent between calendar years.

3. The *Terrorism Component* produces simulations for mortality shocks arising from a terrorist attack based on a multi-level trinomial tree. In each level, there are three possible outcomes with different probabilities: "Failure", i.e. no deaths have occurred; "Success" of the attack, so a random number of deaths within a given range is assumed; or "Escalate", which means that the attack was more severe than attacks corresponding to the current level implying that the model jumps to the next level with a higher range of possible death counts. According to Linfoot (2007), the probabilities and ranges are calibrated to the terrorism model from the *US State Department* (for 1999-2003) and the *National Counterterrorism Center* (for 2004).

[13]Based on Linfoot (2007) and Logisch (2007).

For each basic component 250,000 simulations are produced and combined to estimate annualized as well as cumulative expected losses and several loss probabilities such as the probability that the trigger level of a certain tranche will be reached. Even though the terrorism component only has a marginal impact for these estimations, it was included into Milliman's model since the Vita II deal upon investors' requests. Furthermore, mortality shocks due to natural disasters are not modeled explicitly as *"such events would not have resulted in a large enough number of deaths to cause a loss to any Class of the Notes [within the Tartan deal]."*[14]

As noted above, Milliman's model was the primary basis for investors' and rating agencies' decisions within all CATM transactions thus far since Milliman was appointed as the risk modeling firm in each deal. The reason for this "monopoly-position" is probably that Milliman was the first to have a model for catastrophe mortality risk which was accepted by the market. Since they are part of several ongoing shelf programs and since investors have become accustomed to seeing Milliman involved in the deals, it is very likely that they also will play a dominant role in the future. However, rumors in the market indicate that RMS is currently working on a transaction together with a large American insurer. As RMS reports substantially higher loss probabilities than Milliman, the consequences of an involvement of RMS in future transactions may considerably change current spread levels.

While RMS' idea of building an epidemiological model is interesting, from a mathematical point of view the massive amount of necessary parameters is problematic. For example, when including parameter uncertainty or when conducting sensitivity analyses confidence bands for the loss probabilities become very large; therefore, we will follow Milliman's approach and present an actuarial model in the next subsection. Following the same basic approach, we also keep our results comparable to the risk profiles derived by Milliman, which have been used in the transactions considered in this thesis.

4.2 Our Model for Analyzing and Pricing Mortality Contingent Catastrophe Bonds

In order to analyze and eventually price CATM bonds, it is necessary to model the underlying combined mortality index. While Lin and Cox (2006) and Cox et al. (2006) directly model the weighted population death rate from equation (4.1), we consider modeling the underlying cohort specific death rates. Understanding the index as a function of these death rates is, in our opinion, more adequate as

[14] cf. Linfoot (2007).

the combined index corresponds to the specifications of a certain transaction, and therefore, a "direct" model of the index or the weighted population death rate will naturally be bound to the deal in view. Furthermore, we believe that the more basic approach coheres better with the structure of the problem and permits the consideration of more general transaction structures.

Our model specification belongs to the general setup considered in Subsection 2.8.2, which is in line with Miltersen and Persson (2005), who define the force of mortality based on the so-called intensity based approach from Lando (1998). As previously discussed, in order to guarantee analytical tractability of a model specification, it is sufficient to specify a process $X^1 = (Y^1, Y^2, Z)$ satisfying the conditions stated in Assumption 2.7.1 as well as positive functions $\chi^1 \in \mathbb{R}^r_{0+}$ and $\chi^2(t) \in \mathbb{R}^{1\times d_2}_{0+}$ such that mortality intensities $\lambda_{x_0}(t)$ can be written as

$$\lambda_{x_0}(t) = \lambda(x_t, t) = \lambda(x_t, X^1(t)) = Z(t)\chi^1(x_t) + \chi^2(x_t)Y^2(t)$$

with $x_t = x_0 + t$. Dahl et al. (2006) propose mortality intensities $\lambda(x_t, t)$ of the form

$$\lambda(x_t, X^1(t)) = \chi^2(x_t)\, Y^2(t),$$

where Y^2 is a time in-homogeneous mean reverting square root diffusion process, the parameters of which can also depend on x_0, and the initial mortality intensities χ^2 are of the Gompertz-Makeham form. However, as pointed out by Lin and Cox (2006) "mortality jumps" arising from catastrophes such as pandemics should be taken into account when modeling the evolution of mortality, particularly when focusing on catastrophic mortality risk. Therefore, we extend the model by Dahl et al. (2006) on the one hand by including mortality jumps in the form of a "self-affecting" Ornstein-Uhlenbeck process Y^2_2 but restrict it by choosing a time homogeneous mean reverting square root diffusion Y^2_1 and considering a simpler, pure Gompertz form for the initial mortality intensities. Thus, we choose $Y^2 = (Y^2_1, Y^2_2)$ and propose mortality intensities of the form[15]

$$\lambda(x_t, t) := \lambda(x_t, Y_1(t), Y_2(t)) = Y_1(t)\, e^{b\, x_t + c} + Y_2(t), \tag{4.4}$$

where $b, c \in \mathbb{R}_+$, Y_1 evolves according to the SDE

$$dY_1(t) = \alpha\ (\eta - Y_1(t))\ dt + \sigma\sqrt{Y_1(t)}\, dW(t), \quad Y_1(0) > 0 \tag{4.5}$$

and $Y_2(t)$ is governed by the SDE

$$dY_2(t) = -\kappa Y_2(t)\, dt + dJ(t), \quad Y_2(0) = 0. \tag{4.6}$$

[15]For ease of exposition, we simply write (Y_1, Y_2) instead of (Y^2_1, Y^2_2) in the following.

Here α, η, σ, κ are positive constants, W is a standard one dimensional Brownian motion, and J a point process with intensity

$$\mu(t) = \mu_0 + \xi^{(1)} Y_2(t), \quad \mu_0, \xi^{(1)} > 0,$$

and positive, independent $Exp\left(\frac{1}{\zeta}\right)$-distributed jump sizes. Note that the specification of the jump intensity μ entails a feedback property of mortality jumps. Past jumps make further jumps more likely which allows for modeling a possible "shock wave" behavior of pandemics.

In general, the proposed mortality intensities in (4.4) have an affine structure (cf. Section 2.7) which gives rise to a semi-analytical formula for the survival probabilities implied by them. In the particular nice case of $\mu(t) \equiv \mu_0$, i.e. $\xi^{(1)} = 0$, Y_2 becomes a so-called *Gamma-OU* process (see e.g. Chapter 15 of Cont and Tankov (2004)) yielding the following formula for the survival probabilities:

Proposition 4.2.1 *If mortality intensities $\lambda(x_t, t)$ satisfy (4.4) and Y_1 and Y_2 evolve according to the SDEs (4.5) and (4.6) with $\mu(t) \equiv \mu_0$, survival probabilities are given semi-analytically as*

$$\begin{aligned} {}_{T,t}p_{x_0+t} &= \mathbb{E}\left[\exp\left\{-\int_t^T \lambda\left(x_0+s, Y_1(s), Y_2(s)\right)\, ds\right\}\middle| Y_1(t), Y_2(t)\right] \\ &= \exp\left\{u(T-t) + v(T-t)Y_1(t)e^{b(x_0+t)+c}\right\} \\ &\quad \exp\left\{-\frac{Y_2(t)}{\kappa}\left(1-e^{-\kappa(T-t)}\right) - \frac{\mu_0\zeta(T-t)}{\kappa+\zeta}\right\} \\ &\quad \exp\left\{\frac{\mu_0\zeta}{\kappa+\zeta}\log\left[1+\frac{\zeta}{\kappa}\left(1-e^{-\kappa(T-t)}\right)\right]\right\}, \end{aligned} \tag{4.7}$$

where u and v satisfy the following Riccatti ordinary differential equations (ODEs)[16]

$$\begin{aligned} \dot{u}(x) &= \alpha\eta\, v(x)\, e^{b(x_0+T-x)+c},\ u(0) = 0, \\ \dot{v}(x) &= -1 - (\alpha - b)\, v(x) + \frac{1}{2}\sigma^2 \left(v(x)\right)^2 e^{b(x_0+T-x)+c},\ v(0) = 0. \end{aligned} \tag{4.8}$$

Proof: The result is a special case of Theorem 2.7.1 and the ODEs stated there. Generally, due to their independence Baseline and Catastrophe Component can separately be considered yielding two independent systems of ODEs. The explicit solutions of the ODEs in case of the Catastrophe Component are obtained from a special case of the transform in Appendix A.

□

[16]The ODEs have to be understood as ODEs in x.

Since data on catastrophe mortality events is extremely sparse (see in particular the next section), the specification of Proposition 4.2.1 with constant jump intensity μ_0 will be the one which we consider for the remainder of the chapter. The more general specification would be an alternative if more data was available. Based on Proposition 4.2.1, by solving the ODEs from equation (4.8), we are able to compute survival probabilities and then use the "classical actuarial toolbox" to price life insurance products.

So far, we have fixed a specific probability measure P even though different investors may have different opinions regarding the evolution of the future mortality implying different individual measures. Furthermore, prices for mortality contingent claims may include a risk premium for the inherent risk, which also leads to a different probability measure, the so-called pricing measure Q. Informally, by a change of measure, the structure of the model can be altered considerably: Not only could it affect the intensity process λ, but a market price for unsystematic or idiosyncratic risk may be included (see e.g. Biffis et al. (2005)). We will not consider loadings for the idiosyncratic component as we regard a whole population. However, when limiting the perspective to an insurer's portfolio, there may (or may not) be reasons for a premium for unsystematic risk (see Subsection 2.8.2 and Bauer and Russ (2006) for a discussion of this issue). Furthermore, when allowing for an (almost) arbitrary change of measure the structure of the force of mortality could change tremendously under the "new" measure. Therefore, similarly to Dahl et al. (2006) we restrict ourselves to choices where the parameters of our setup can be changed but not the process' structure.

From a mathematical point of view, the analytical tractability of the presented model is a valuable and important feature that can be considered as a first advantage. A detailed discussion of our model choice in the context of actual mortality data is postponed to the next section where different calibration procedures are presented.

4.3 Calibration of the Model

When applying a financial model to determine risk-measures such as loss probabilities or expected losses, historical data is usually used to determine a parametrization which matches the past experience. However, as we are considering the evolution of mortality not only past experiences but also the particular properties of mortality have to be taken into account. For example, mortality rates are positive. Furthermore, the projection of future death rates is not a purely statistical problem as demographic considerations should also be taken into account. Therefore, when calibrating our model to historic data in Subsection 4.3.1, we also incorporate de-

mographical aspects.

When pricing contingent claims, on the other hand, it is usually not sufficient to rely on historical data as prices include premiums for the adopted risk. This risk premium generally cannot be determined endogenously, but it is implied by the market. Aside from catastrophe mortality securitization transactions, life insurance prices are subject to catastrophic mortality risk. Hence, a natural approach for determining a risk-adjusted parametrization for our model is to extract it from insurance prices. This indirect approach is presented in Subsection 4.3.2, whereas in Subsection 4.3.3 we calibrate the model using data directly from CATM securitization transactions.

4.3.1 Backtesting the Model and Historical Parametrizations

Most CATM transactions to date were primarily exposed to US mortality experience. Therefore, we limit our considerations to American mortality data. Furthermore, we focus on male mortality experience because male death rates are usually weighted more heavily than female death rates within the combined mortality indices (see Table 4.1); however, when assuming independence of the "regular" mortality evolution for the different cohorts and a simultaneous occurrence of pandemics in the relevant countries as in the Milliman model, including female or non-US mortality data is straight-forward. Without these rather rigorous assumptions, i.e. when allowing for correlations and diversification effects across genders and populations, the calibration procedure will get more complex as correlations need to be estimated and incorporated into the model. We leave the exploration of this issue for future work.

Our model consists of two independent components: A diffusion part which models the "regular" evolution of mortality, i.e. when no catastrophic event occurs, and a jump-part which models pandemics and other catastrophes. In what follows, we will refer to these two components as the *Baseline Component* and the *Catastrophe Component*, respectively. The independence of the two components allows us to carry out the calibration procedures separately.

The Baseline Component[17]

We use annual, periodic male mortality data as available from the Human Mortality Database for our considerations.[18] There, age specific death rates $m(x,t)$ are available for each year t from 1959 until 2003. From these death rates, we derive the sample mortality intensities $\tilde{\lambda}$ by the following approximation (for a proper definition and properties of death rates $m(x,t)$ see e.g. Bowers et al. (1997)):

$$\begin{aligned}\tilde{\lambda}(x_t+0.5,t+0.5) &= \tilde{\lambda}(x_t+0.5,t+0.5)\frac{\int_0^1 l(x_t+s,t+s)\,ds}{\int_0^1 l(x_t+s,t+s)\,ds}\\ &\approx \frac{\int_0^1 l(x_t+s,t+s)\,\lambda(x_t+s,t+s)\,ds}{\int_0^1 l(x_t+s,t+s)\,ds}\\ &= m(x_t,t), \end{aligned} \tag{4.9}$$

where $l(x_t,t)$ denotes the exposures, i.e. the number of individuals within the relevant cohort of x_t-year aged males at time t. In order to simplify notation, we set the inception date $t=0$ to mid 1959, i.e. 30.06.1959, and thus we are given the spot intensities $\tilde{\lambda}(x_t,t)$ for ages $x_t \in \{0.5, 1.5, ..., 100.5\}$ in years $t=0$ (1959.5) through $t=44$ (2003.5).[19] Within our model, neglecting the influence of the catastrophe component, the endogenous intensities $\lambda(x_t,t)$ are of the form (see equation (4.4))

$$\lambda(x_t,t) = Y_1(t)\,e^{b\,x_t+c}, \tag{4.10}$$

where we conveniently set $Y_1(0)=1$. Therefore, $\lambda(x_0,0) = e^{b\,x_0+c}$ is simply given by the Gompertz form, and the parameters b as well as c can be determined by an exponential regression on $\tilde{\lambda}(x_0,0)$; we obtain $b=0.08117916$ and $c=-8.7674591$. The comparison of the "actual" and model-endogenous mortality intensities is displayed in Figure 4.4. We find that the Gompertz approximation fits the data quite well. Particularly for years below 85, which are most relevant for the calculation of the combined mortality index (cf. Table 4.1), we only see slight deviations.

Similarly, Gompertz forms can be derived for all years t. However, within our model, b and c are constant over the years. By the relationship

$$\lambda(x_t,t) = e^{b\,x_t+c+\log\{Y_1(t)\}} \Leftrightarrow \log\{\lambda(x_t,t)\} - b\,x_t - c = \log\{Y_1(t)\},$$

[17]The Baseline Component considered here should not be confused with the Baseline Component introduced in Section 3.1. In Section 3.1, the Baseline Component modeled an observed deterministic intra-month default pattern, while here the "regular" yet stochastic evolution of mortality is modeled.

[18]Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at *www.mortality.org* or *www.humanmortality.de* (downloaded 11/03/2006 (1959-2002) and 04/11/2007 (2003)).

[19]Since the mortality index is only slightly affected by ages less than 30 and since there are structural deviations for very young ages and ages around 20 (so-called mortality humps), we only consider ages above 30 for the calibration.

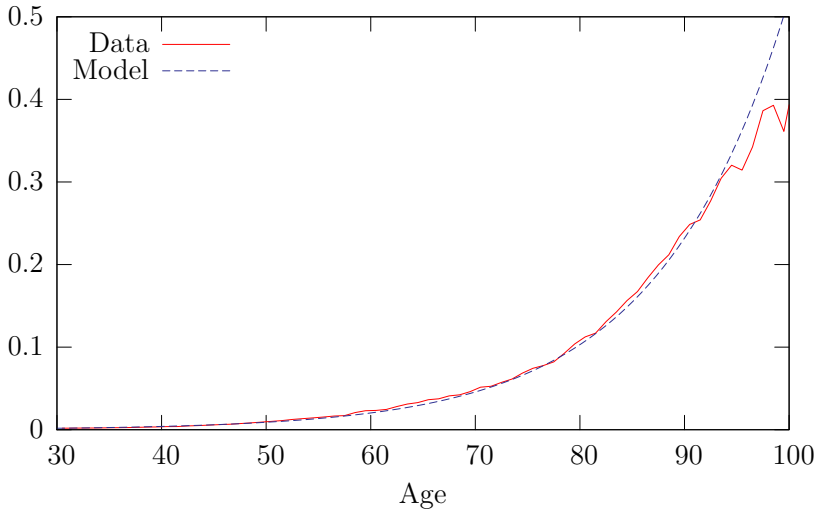


Figure 4.4: Original and calibrated mortality intensities for year 0 (1959).

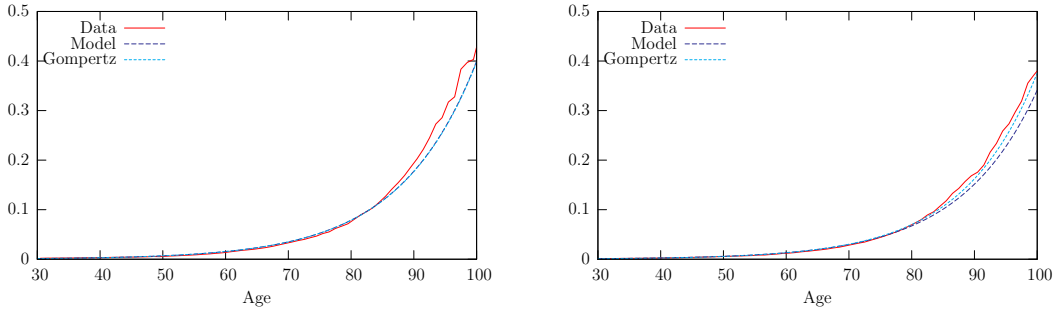


Figure 4.5: Original ("Data") and calibrated ("Model") mortality intensities for years 36 (1995) and 43 (2002). Estimations of Gompertz mortality intensities which are only based on the respective year are displayed, too.

we can estimate $\log\{Y_1(t)\}$ as the mean of $\log\left\{\tilde{\lambda}(x_t,t)\right\} - b\,x_t - c$, $x_t = 0.5,\, 1.5, \ldots$ and, hence, our model $\lambda(x_t,t)$ by equation (4.10). For example, in Figure 4.5 the actual-, the model-, and the Gompertz-mortality intensities for the years 36 (1995.5) and 43 (2002.5) are shown. Again, we can see that the model and the Gompertz intensities fit the data well in the relevant ages. However, the growth of the actual intensities is super-exponential for older ages, i.e. while the curve gets less steep for younger ages, it increases very fast for older ages. Furthermore, we find that the model and the Gompertz intensities are very close; for year 36 deviations are hardly noticeable, and for year 43 the deviations in the more relevant ages below 85 are also rather small.

By this procedure, we obtain a time series of the $Y_1(t)$ for $t = 0, \ldots, 43$, shown in Figure 4.6, which can be used to estimate the parameters of the square root diffusion

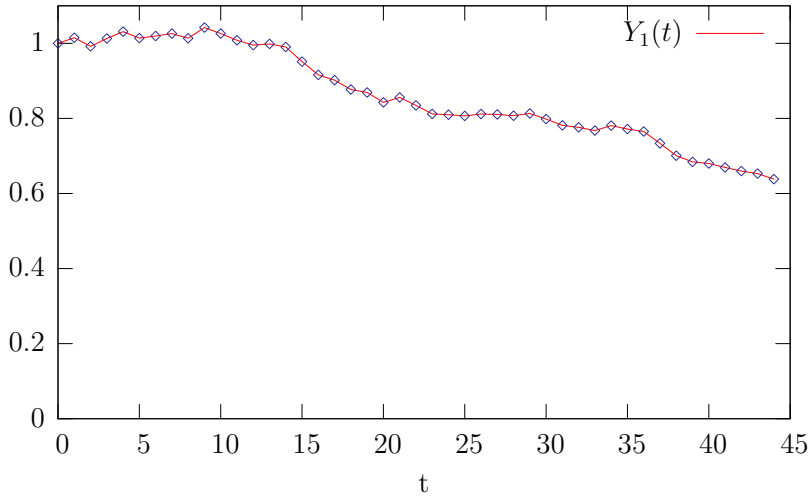


Figure 4.6: Estimated time series of the Baseline Component Y_1 from year 0 (1959) to 44 (2003).

process α, η, and σ by estimators for the Cox-Ingersoll-Ross (CIR) interest rate model (see Cox et al. (1985)). However, there are two potential pitfalls.

Since we want to separate the influences of catastrophic events, such as pandemics, and "regular" deviations in mortality, we should not include peaks which are due to such catastrophic events. In the time between 1959 and 2003, there were two major occurrences: the so-called Hong-Kong Flu from 1968 and the so-called Russian Flu from 1977. The first is quite noticeable in our time series as there is a peak at the data point $t = 9\,(1968.5)$. In order to disregard these influences, we smooth the time series by linearly interpolating these data points by the surrounding ones and taking the interpolated value instead of the recorded one whenever the interpolated value is lower.

We cannot observe a mean reversion trend in the time series. This fact could be interpreted as a problem with the specification or the general structure of the model, i.e. that a mean reverting process does not present a suitable choice. In fact, the question whether mean reverting processes are adequate for describing the evolution of mortality has been raised before in the literature (see e.g. Luciano and Vigna (2005)).

In this regard, Oeppen and Vaupel (2002) show that the average life expectancy in the country with the current highest life expectancy has increased almost linearly by slightly less than three months per year over the last 160 years. Even though the annual records were set by only two nations since 1975, namely Iceland and Japan, the observations for other industrialized countries are quite similar. Therefore,

they do not believe in a barrier for the life expectancy, which would be implied by a positive mean reversion level. Furthermore, they note that there is a long history of conjectured barriers, which were all broken only shortly after their publication.

A continuing linear trend of life expectancies may even require mortality intensities to decrease faster than observed over recent years: Keyfitz (1985) shows that a reduction of mortality intensities across all ages of $\delta\,\%$ would imply an increase of the life expectancy by $\delta\, H(t)\,\%$, where $H(t)$ is the so-called *demographic entropy* at time t, and investigations by Olivieri (2001) and Pitacco (2004) indicate that the demographic entropy is decreasing towards zero.[20]

However, Olshansky et al. (2001) believe that a faster decrease of death rates is very improbable. They argue that, on the one hand, death rates for young ages are almost at their minimum value.[21] Therefore, reductions in high ages would need to account for increasing life expectancies, which eventually would lead to a so-called negligible senescence, i.e. mortality rates would remain constant for all attainable ages and aging would not reduce survival probabilities. According to the authors, this contradicts basic biological ideas and is thus not likely, if not impossible. On the other hand, they note that social-political and economic reasons indicate that life expectancies will not increase continuously.

We do not want to join this discussion (see Kristen (2007) for more details). All in all, there is no general agreement among demographers regarding the future evolution of life expectancies.

Our model does not allow for systematically faster decreasing mortality rates in the future when choosing a positive mean reversion level η and speed of mean reversion α.[22] Furthermore, when applying the maximum likelihood estimators from Walter (1996) these result in a negative speed of mean reversion α and a mean reversion level η greater than 1, which indicates a strictly negative drift term of the process displayed in Figure 4.6. We may solve this problem by allowing for a deterministic function η_t as the mean reversion level in our specification of Y_1 (see equation (4.5)) rather than a constant. For example, η could be replaced by

$$\eta(t) = e^{-\eta_1 t} + \eta_2, \quad \eta_1,\eta_2 > 0.$$

[20] In the literature, this development is usually referred to as *rectangularization* as the shape of the mortality intensity curve gets more "rectangular", i.e. deaths occurrence is concentrated around a certain modal age (see also Figure 4.5).

[21] Demographers generally assume a minimal level of death rates for all age groups, which is motivated by the natural occurrence of accidents etc.

[22] This is due to the fact that in our specification, the drift of Y_1 decreases as it gets closer to the mean reversion level.

A deterministic mean reversion level $\eta(t)$ would not affect the analytical tractability of the proposed model. However, the number of model parameters would increase complicating calibration procedures. This particularly seems problematic in view of the limited data availability of only 44 data points. Moreover, even with a constant mean reversion level, our specification permits modeling all presented demographic viewpoints by choosing a coherent value of η. Therefore, we restrict ourselves to constant mean reversion levels; exploring the possibility of choosing a deterministic mean reversion level would be an interesting topic of future research.

Instead of following only one particular demographic opinion, we will consider three parametrizations P1, P2, and P3, where the first ($\eta = 0.6$) and the third ($\eta = 0$) correspond to the "extreme" points of view from Olshansky et al. (2001) and Oeppen and Vaupel (2002), respectively, whereas P2 ($\eta = 0.4$) is a choice in the middle. Keeping η fixed, we can obtain η-"adjusted" estimators for α and σ as follows: Discretizing the SDE for Y_1 from equation (4.5) yields

$$Y_1(t+1) - Y_1(t) = \alpha\left(\eta - Y_1(t)\right) + \sigma\sqrt{Y_1(t)}N(t),$$

where $N(t)$, $t = 0, ..., T-1 = 43$ are independent $N(0,1)$-distributed random variables. Therefore,

$$\sigma N(t) = \frac{Y_1(t+1)}{\sqrt{Y_1(t)}} - \sqrt{Y_1(t)} + \alpha\sqrt{Y_1(t)} - \frac{\alpha\eta}{\sqrt{Y_1(t)}},$$

and the standard estimators for the mean and the variance (ideally) give

$$\begin{aligned} \frac{1}{T}\sum_{t=0}^{T-1}\frac{Y_1(t+1)}{\sqrt{Y_1(t)}} - \sqrt{Y_1(t)} + \alpha\sqrt{Y_1(t)} - \frac{\alpha\eta}{\sqrt{Y_1(t)}} &\stackrel{!}{=} 0 \quad \text{and} \\ \frac{1}{T-1}\sum_{t=0}^{T-1}\left(\frac{Y_1(t+1)}{\sqrt{Y_1(t)}} - \sqrt{Y_1(t)} + \alpha\sqrt{Y_1(t)} - \frac{\alpha\eta}{\sqrt{Y_1(t)}}\right)^2 &\stackrel{!}{=} \sigma^2, \end{aligned}$$

respectively. As η is given, solving the first equation for α yields the estimator

$$\hat{\alpha} = \frac{\sum_{t=0}^{T-1}\frac{Y_1(t+1)}{\sqrt{Y_1(t)}} - \sqrt{Y_1(t)}}{\sum_{t=0}^{T-1}\frac{\eta}{\sqrt{Y_1(t)}} - \sqrt{Y_1(t)}},$$

and plugging $\hat{\alpha}$ back into the second equation gives the estimator for σ,

$$\hat{\sigma} = \sqrt{\frac{1}{T-1}\sum_{t=0}^{T-1}\left(\frac{Y_1(t+1)}{\sqrt{Y_1(t)}} - \sqrt{Y_1(t)} + \hat{\alpha}\sqrt{Y_1(t)} - \frac{\hat{\alpha}\eta}{\sqrt{Y_1(t)}}\right)^2}.$$

Table 4.4 displays the resulting parametrizations.

	Fixed	Estimated	
	η	α	σ
P1	0.6	0.0325	0.01647
P2	0.4	0.01829	0.01582
P3	0.0	0.00976	0.01552

Table 4.4: Estimated parametrizations of the Baseline Component Y_1 for fixed η.

The Catastrophe Component

In Lin and Cox (2006), catastrophic events are modeled as multiplicative shocks on the combined mortality index. While this approach could be reasoned by the idea that due to weaker immune systems the elderly population may be most affected by a possible pandemic in absolute terms, cytokine storms, which are believed to have been present during severe pandemics such as the Spanish Flu, may lead to a disproportionate amount of deaths in younger ages (cf. Section 4.1.3). Furthermore, man made catastrophic events such as severe terror attacks are not likely to affect older aged individuals more than younger individuals. Therefore, it seems to be appropriate to model catastrophic events as additive shocks, and we included the catastrophe component as an additive jump part to the mortality intensities (see equation (4.4)).

For the calibration of the catastrophe component, we rely on the data from Linfoot (2007), where the frequency and severity of historical occurrences of infectious disease epidemics based on U.S. population experience are provided. This choice is motivated by the fact that this data was used by the risk modeling firm Milliman within the Tartan transaction, and we want to keep our findings comparable.

Linfoot (2007) reports an annual frequency of 7.4% (31 occurrences in the past 420 years) and severities for 5 (6) specific occurrences (model points)[23]. Therefore, we set the jump intensity of the compound Poisson process within the catastrophe component to $\mu_0 = 7.4\%$. Percentiles for the likelihood of each model point are derived by considering the number of equal or worse pandemics in relation to all occurrences. For example, the Spanish Flu is taken as a 1 in 420 years event, and given the annual frequency of 7.4%, this yields the $\frac{1}{420}\frac{1}{0.074} \approx 0.032$ percentile.

The impact on the force of mortality which a catastrophic event implies τ years after its first occurrence in our model is given by $\Delta e^{-\kappa\tau}$ with Δ denoting the initial "jump". We assume that the event basically affects the mortality for one year only,

[23] 1918 (Spanish Flu), 1957 (Asian Flu), 1968 (Hongkong Flu), 1977 (Russian Flu) and 2003 (SARS). We omitted one data point called "Adjusted 1918-20" as it was fixed at the 0.0 Percentile Level.

i.e. that the influence after a year has decreased to only 1% of the initial impact, and thus set

$$\Delta e^{-\kappa \cdot 1} = 0.01\Delta \Longleftrightarrow \kappa = 4.6052.$$

For fitting the exponential distribution of Δ, the given severities for the five model points are used. However, there the (multiplicative) excess mortality as a percentage of the mortality probability over all ages is provided. In order to use them for our considerations, they need to be "translated" to additive excess mortality intensities. As the population is not homogeneous across all ages, it is not sufficient to compute excess mortality intensities resulting from distorted mortality rates for every age and take the mean, but they have to be properly weighted: For the male population death rate at time t, we have

$$\tilde{q}_t^m = \sum_{\text{all ages } x} \tilde{\omega}_x^{(t)}\, \hat{q}_{m,x,t},$$

where $\tilde{\omega}_x^{(t)}$ are the male population weights, and $\hat{q}_{m,x,t}$ is the mortality rate for an x-year old male at time t.[24] Hence, if ε denotes the excess mortality rate, the corresponding initial impact Δ of the catastrophic occurrence can be approximated by

$$\begin{aligned}
(1+\varepsilon)\tilde{q}_t^m &= (1+\varepsilon) \sum_{\text{all ages } x} \tilde{\omega}_x^{(t)}\, \hat{q}_{m,x,t} \\
&\stackrel{!}{=} \sum_{\text{all ages } x} \tilde{\omega}_x^{(t)} \left(1 - e^{-\int_0^1 \lambda(x+s,t+s)+\Delta e^{-\kappa s}\, ds}\right) \\
\Rightarrow \frac{\Delta}{\kappa}\left(1-e^{-\kappa}\right) &= \log\left\{\frac{1-\tilde{q}_t^m}{1-(1+\varepsilon)\tilde{q}_t^m}\right\}. \qquad (4.11)
\end{aligned}$$

Therefore, given ε, we still need to fix mortality rates and population weights to determine Δ. We use mortality rates as implied by the Gompertz forms and population weights as provided by the *U.S. Census Bureau*[25] for years 1959 and 2003, respectively. By matching the quantiles of an exponential distribution for the model points, we arrive at the parameters displayed in Table 4.5. Aside from an exponential distribution for the jumps, we also fit a Gamma distribution. Keeping the mean at the same level as for the exponential distribution in order to obtain comparable results and choosing the parameters that provide the best match with the model points in a (weighted) Least Squares sense, we arrive at the parameters

[24] Note that the weights $\tilde{\omega}_x^{(t)}$ are the actual population weights and, thus, do not coincide with the deal-specific weights from equation (4.1).

[25] U.S. Census Bureau. National population estimates (male), *www.census.gov/popest* (downloaded 04/19/2007).

		1959	2003
Exp($\frac{1}{\zeta}$) *Distribution*	ζ	0.003147	0.002799
Gamma(g_1,g_2) *Distribution*	g_1	0.57298	0.57293
	g_2	182.08	204.71

Table 4.5: Estimated parametrizations for the jump size distributions.

displayed in Table 4.5.

We find that for a different demographic structure and different mortality rates, the calibrated parameters are quite different: For the exponential distribution, the mean is reduced by approximately 11% from the 1959 to the 2003 estimates. However, it is not clear which parametrization is more adequate; on the one hand, the 2003 demographic structure resembles the actual structure today, but on the other hand, the 1959 demographic structure may better cohere with the demographic structure when (severe) pandemics occurred. In order to keep our presentation concise, if not stated otherwise, we rely on the parametrization for 2003 and exponentially distributed jumps.

It is worth noting that due to the data used, the catastrophic component is subject to a high parameter uncertainty. For example, when only considering pandemic data from the last century, we may still use the data points from Linfoot (2007), but the resulting quantiles change considerably: When proceeding analogously, for the Spanish Flu we would obtain the $\frac{1}{100}\frac{1}{0.074} \approx 0.135 > 0.032$ percentile. Moreover, the given annual frequency of 7.4% is rather high in comparison to the frequency proposed by RMS ($3-4\%$) or by Cox et al. (2006) ($\approx 3.3\%$). Thus, it is necessary to conduct detailed sensitivity analyses for the catastrophe component. Furthermore, man made catastrophes are not explicitly modeled because we assume that the structural affect on mortality rates is very similar to a pandemic occurrence. However, as past events had a negligible effect on general population mortality they are not considered in our calibration procedure.

4.3.2 Risk-Adjusted Calibration Based on Insurance Prices

If life insurers knew the future evolution of mortality, mortality risk management would be simple: With an increasing number of insured, the risk per sold policy would decrease to zero by the Law of Large Numbers. However, aside from this diversifiable, "unsystematic" mortality risk insurance companies are also exposed to "systematic" mortality risk as the future evolution of aggregate mortality actually

is not deterministic. Catastrophe mortality transactions allow insurers to transfer a part of this systematic risk, namely the part due to the possible occurrence of a catastrophic event, to the capital market. These transactions or, more specifically, the excess spreads are financed by the insurer and, hence, eventually by insurance premiums. This means that these premiums must account for catastrophic mortality risk. Thus, it is a natural idea to use life insurance prices to derive a risk adjusted parametrization for the mortality intensity process.

Similar ideas have been proposed for pricing longevity bonds (see Lin and Cox (2005)) and classical catastrophe derivatives (see Muermann (2003)). As pointed out by Bauer and Russ (2006), some conditions need to be satisfied regarding insurance prices and the insurance market in order to derive the risk premium for systematic mortality risk from insurance prices. In particular, insurance prices should not include a loading for unsystematic mortality risk,[26] and the insurance market should be free of arbitrage. We refer to their article for a discussion of these assumptions.

The basic approach is straight-forward: Using our model, we compute prices for term life insurance policies and derive parameters such that the model-endogenous prices match market quotes optimally in the least squares sense. We consider term life insurance contracts with different maturities T_i and individuals of different ages x_0 at inception,[27] who get paid a fixed death benefit D against fixed, monthly premiums P. Hence, the expected discounted value of the benefits $\mathrm{EB}_{T_i,x_0}(\mathrm{D})$ and the premiums $\mathrm{PB}_{T_i,x_0}(\mathrm{P})$ are given by the following equations:

$$\begin{aligned}
\mathrm{EB}_{T_i,x_0}(\mathrm{D}) &= \mathbb{E}\left[\sum_{t=0}^{T_i 12-1} \mathrm{D} e^{-r\frac{t+1}{12}} 1_{\frac{t}{12}<\tau_{x_0}\leq\frac{t+1}{12}}\right] \\
&= \sum_{t=0}^{T_i 12-1} \mathrm{D} e^{-r\frac{t+1}{12}} \left(\mathbb{E}\left[1_{\tau_{x_0}>\frac{t}{12}}\right] - \mathbb{E}\left[1_{\tau_{x_0}>\frac{t+1}{12}}\right]\right) \\
&= \mathrm{D}\sum_{t=0}^{T_i 12-1} e^{-r\frac{t+1}{12}} \left({}_{\frac{t}{12}}p_{x_0} - {}_{\frac{t+1}{12}}p_{x_0}\right), \\
\mathrm{PB}_{T_i,x_0}(\mathrm{P}) &= \mathbb{E}\left[\sum_{t=0}^{T_i 12-1} \mathrm{P} e^{-r\frac{t}{12}} 1_{\tau_{x_0}>\frac{t}{12}}\right] \\
&= \mathrm{P}\sum_{t=0}^{T_i 12-1} e^{-r\frac{t}{12}} {}_{\frac{t}{12}}p_{x_0},
\end{aligned}$$

[26] Technically, this means that the change of measure implied by including a risk premium should not affect the idiosyncratic jump component (cf. Section 4.2).

[27] Note that in contrast to the last subsection, for the remainder of the text we set the inception date 0 to January 1st, 2006.

where r denotes the (constant) rate of interest. For the remainder of the chapter, we assume a constant short rate of $r = 5.05\%$, which corresponds to the 1 year U.S. treasury constant maturity date rate from 01/11/2007.[28] For a given set of parameters, the quantities ${}_{t,0}p_{x_0}$ can be conveniently calculated using equation (4.7). By the actuarial principle of equivalence (see e.g. Bowers et al. (1997)), we then obtain the model-endogenous premiums

$$\hat{\mathrm{P}}_{T_i,x_0} = \frac{\mathrm{EB}_{T_i,x_0}(\mathrm{D})}{\mathrm{PB}_{T_i,x_0}(1)},$$

which depend on the parameter choice. Hence, the task is to find parameters such that the target function

$$\sum_{T_i,x_0} \left(\hat{\mathrm{P}}_{T_i,x_0} - \mathrm{P}_{T_i,x_0}\right)^2 \to \min$$

is minimized, where P_{T_i,x_0} are the actual market quotes. Our data set contains prices as provided by *Quickquote.com* for male Californians[29] with ages ranging from 25 to 55 and maturities from 10 up to 30 years. This calibration routine as well as all other numerical calculations are implemented in *C++* using routines from the *GNU Scientific Library* (GSL); in particular, we make use of an ODE solver provided in the GSL based on the *Runge-Kutta* method.

But again, there are two potential pitfalls:

1. Death rates for the population of insured and the general population differ considerably, but CATM bond prices depend on population mortalities, whereas insurance prices depend on insured mortalities. Hence, assuming equality would not be adequate. However, this is not the case for mortality improvements: Even though mortality improvements were somewhat higher for the population of insured in comparison to the general population, the deviation is rather small. For example, for the incorporation of mortality trends into the *Valuation Basic Table 2001* (VBT 2001) the American Academy of Actuaries reports that they relied on the improvements for the general population (cf. American Academy of Actuaries (2002)). Therefore, we assume that mortality improvements, which in our model are governed by the process Y_1, are alike for the population of insured and the general population. Furthermore, it seems to be reasonable that the exposure to catastrophes do not differ between the two populations, and therefore we also assume that

[28] FRED (Federal Reserve Economic Data) provided by the St. Louis Federal Reserve Bank (*www.research.stlouisfed.org*). As the yield curve has only a mild downward slope, we consider the one year treasury yield as a fair proxy for our considerations.

[29] Non-Smoking, Standard-Plus, monthly premium payments, $100,000 coverage.

the process Y_2 is the same for both populations. Thus, it is sufficient to apply different parametrizations for the initial mortality intensity e^{bx_0+c} to the different populations.

2. Insurance prices include adjustments for selection effects, which e.g. arise due to mandatory health examinations before underwriting the policies. Hence, when determining insurance prices based on mortality rates without selection effects included the resulting prices are higher than corresponding prices based on selection tables. Since our model does not take selection effects into account, resulting parametrizations tend to underestimate risk and respective premiums included in insurance prices but yield a lower bound.

In order to carry out the optimization algorithm, we still need to fix the initial mortality intensity, i.e. we need to find appropriate parameters b and c. We use the VBT 2001, where period (spot) mortality rates $q_x^{(2001)}$ are provided. By a similar approximation to the one used in equation (4.9), we can derive mortality intensities via the relationship

$$\begin{aligned} q_{x_t}^{(2001)} &= 1 - e^{-\int_0^1 \tilde{\lambda}(x_t+s,2001+s)\,ds} \\ &\approx 1 - e^{-\tilde{\lambda}(x_t+0.5,2001+0.5)}. \end{aligned} \tag{4.12}$$

However, mortality intensities for $t = 0$, i.e. $2006+0$ rather than for $2001+0.5$, are needed. We approximate them using the same methodology as was implemented for the derivation of the VBT 2001 (see American Academy of Actuaries (2002)); here, the 2001 data was extrapolated from 1990-1995 Basic Tables derived by the Society of Actuaries' (SOA) Individual Experience Committee. By applying the same trends, we further extrapolate the 2001 data to 2005.5 and derive mortality intensities $\tilde{\lambda}(x, 2006)$ by relationship (4.12), and hence, parameter $b = 0.09697$ and parameter $c = -10.62217$ by exponential regression. Of course, when choosing the starting value of the continuous part Y_0 different than 1, c has to be adjusted accordingly.

In the calibration procedure, we encounter numerical instabilities due to local minima of the target function. As usual, we solve this problem by considering a large set of different starting values for the optimization algorithm and choose the parameters which imply the minimum value for the target function within the considered set. The resulting parameter estimates are displayed in Table 4.6.[30]

We find that the parametrization for the baseline component significantly differs from the results in the foregoing section. This may be a sign that the death rates

[30] $Y_1(0)$ was fixed at the same level as for the 2006 population mortality (cf. Section 4.3.3), and c was adjusted accordingly.

Fixed Parameters

b	c	$Y_1(0)$	$Y_2(0)$
0.09697472	-10.16051938	0.630240494	0

Estimated Parameters

α	η	σ	κ	ζ	μ_0
0.1528	0.2234	0.0003393	6.370e-16	0.002860	0.06441

Table 4.6: Estimated parameters based on insurance prices.

derived from the VBT 2001 do not present a very good match to the death rates underlying the population of insured in view: The high mean reversion speed as well as the low volatility imply that the spot mortality intensities decrease very rapidly to levels close to ηe^{bx+c}. For the catastrophe component, on the other hand, the resulting expected jump size ζ (jump intensity μ_0) is only slightly increased (decreased) compared to the historical value, but the impact over time, which the occurrence of a pandemic has on mortality rates, is substantially higher since κ is almost zero.

These outcomes indicate that we have to be careful with definite conclusions since we did not include selection effects and since we can not be sure whether the initial mortality data from the VBT 2001 presents a good approximation for the population of insured in view. Considering other mortality tables or a different set of insurance quotes may yield more reliable results. For example, using prices from continuing options for existing term life insurance contracts, where selection effects usually are less dominant, may be worthwhile.

4.3.3 Parameters Implied by Market Prices

When pricing credit derivatives, the parameters of a given model are usually calibrated to market prices of certain securities such as Credit Default Swaps (CDSs) or CDOs with different maturities. Similarly, we may also parametrize our model based on prices of different CATM transactions or tranches within one transaction. In order to derive such an implied parametrization, but also to eventually analyze and price a transaction given some parameters, it is necessary to model the combined mortality index contingent on the basic quantities within our model, i.e. mortality intensities.

As noted earlier in this section, we limit our considerations to one gender and one population only. Actual transactions were based on both genders, but differences

in the evolution of male and female death rates are rather small, particularly in view of catastrophic events. This may not be the case for different populations as a possible epidemic or a severe terrorist attack may occur locally. Into a multidimensional version of our model, a complex dependence structure between the catastrophe components can be incorporated. However, for the baseline component restrictions need to be imposed; for example, independently or fully correlated evolving baseline components will sustain the model's analytical tractability, but empirical investigations are necessary to support such assumptions. Furthermore, the number of parameters will naturally increase.

As mentioned earlier in the text, we primarily focus on the Tartan transaction, which is solely based on US mortality experience. Denoting $\frac{1}{2}\left(\hat{q}_{2005}+\hat{q}_{2004}\right)$ by i_0, the combined mortality index i_t at times $t=2(2007),3(2008)$ in terms of our model is given by (cf. equation (4.2))[31]

$$\begin{aligned} i_t &= \frac{1}{2i_0}\left(\hat{q}_t+q_{t-1}\right) \\ &= \frac{1}{2i_0}\sum_x \omega_{x,m}\left(\left(1-e^{-\int_0^1 \lambda(x+s,t-1+s)\,ds}\right)+\left(1-e^{-\int_0^1 \lambda(x+s,t+s)\,ds}\right)\right) \\ &= \frac{1}{i_0}-\sum_x \frac{\omega_{x,m}}{2i_0}\left(e^{-\int_0^1 \lambda(x+s,t-1+s)\,ds}+e^{-\int_0^1 \lambda(x+s,t+s)\,ds}\right) \end{aligned} \tag{4.13}$$

In order to determine "the value", i.e. the expected discounted payoff under a certain model parametrization, we need to determine the cash flows of the security. As explained in Section 4.1.1, the investor is entitled to coupon payments. During the first measurement period, that is the time before the index is calculated for the first time (here $t=2$), the coupon payments are not at risk. For the remaining time, interest is only paid on the remaining principal, which is determined according to equation (4.3). At maturity, interest for the last period and the remaining principal are disbursed. Table 4.7 shows the resulting cash flows for a nominal of 1 and spread s within the Tartan transaction. To simplify notation, annual instead of quarterly coupon payments are assumed; a and d denote the attachment and detachment point, respectively. Thus, it is sufficient to determine the (joint) distributions of i_2 and i_3 to derive the value as the sum of the expected discounted cash flows.

While there is an approximative method to solve the valuation problem semi-analytically, the derivation is computationally involved as it requires numerical methods for inverting multidimensional Fourier/Laplace transforms and the numerical computation of multidimensional integrals (see the Appendix for details).

[31]Note that in comparison to Table 4.1, the weights $\omega_{x,m}$ need to be adjusted as we consider only male and single age rather than age group weightings.

	Cash Flows
$t = 0$	-1
$t = 1$	$(\text{LIBOR} + s)1$
$t = 2$	$(\text{LIBOR} + s)1$
$t = 3$	$(\text{LIBOR} + s)\left(1 - \min\left\{1, \max\left\{\frac{i_2 - a}{d-a}, 0\right\}\right\}\right)$
	$+\left(1 - \min\left\{1, \max\left\{\frac{i_2 - a}{d-a}, \frac{i_3 - a}{d-a}, 0\right\}\right\}\right)$

Table 4.7: Simplified (yearly instead of quarterly) cash flow scheme for the Tartan bonds.

But aside from the computational difficulties, we are faced with a practical constraint: There are only very few market prices available. Naturally, we need at least as many quotes as there are parameters to be calibrated. In particular, for the Tartan deal there is only a single price available as there is only one (unwrapped) tranche which enables us to only calibrate one parameter implicitly. While we may additionally consider prices of other transactions, as for example the OSIRIS or the Vita III deals, within these securities, the combined mortality index is subject to several different populations (e.g. 60% France for OSIRIS), and assuming the same evolution for the underlying mortality would be rather harsh without empirical investigations.

Therefore, we limit our considerations to the single Tartan tranche. For both implicitly calibrating the parameter(s) as well as analyzing and pricing the contracts, we rely on Monte Carlo methods by simulating the index. This is computationally procurable for only one free parameter and has the additional advantage that we can carry out the computation "exactly" rather than using an approximation. However, when there are more prices available Monte Carlo simulations may not present a feasible choice for calibrating several parameters at a time and the approximative method from the Appendix may be a valuable alternative.

In the calibration procedure, we are left with the choice of which parameters to fix and which parameter to keep variable since only one free parameter may be included. We choose the expected jump size: On the one hand, the baseline component, by definition, does not reflect the attitude towards catastrophic events, which are most important for the transactions, and is, therefore, set to the conservative parametrization P1 (see Table 4.4). On the other hand, as we want to interpret a jump of the catastrophe component solely as a catastrophic event, relatively high values for the jump intensity μ_0 or very low values for κ are not preferable. Hence, we fix μ_0 and κ to the parameter values 0.074 and 4.6052 which were estimated in Subsection 4.3.1. Similarly as for pricing insurance contracts, we further need to fix a parametrization for the initial mortality intensity e^{bx+c}.

	Cl. B Tranche(110%-115%)			*Cl. A Tranche(115%-120%)*		
Scenario	PD(%)	EL(%)	Spread(bps)	PD(%)	EL (%)	Spread(bps)
P1	2.4856	1.5675	61.80	0.9162	0.5789	22.65
P2	2.0954	1.3223	52.53	0.7710	0.4856	19.20
P3	1.8896	1.1908	47.70	0.6844	0.4346	17.29

Table 4.8: Influence of the Baseline Component on the risk profile of the Tartan tranches. Calculated PDs, ELs and tranche spreads are based on the estimated parameters in scenarios P1, P2 and P3. P1 assumes only very mild mortality improvements in the future, P2 medium improvements and P3 assumes strong mortality improvements.

We proceed analogously to the last subsection, i.e we derive $\lambda(x_0, 0)$ based on the given 2003 population mortality data and the projection method from the American Academy of Actuaries (2002). Now, given the spread level of 300bps for the Class B notes of the Tartan deal, we choose ζ such that we fit the price – we obtain an expected jump size of $\zeta = 0.0075726$.

Based on the different parameter choices that we derived in this section, we are now able to price and analyze the CATM bonds from different perspectives.

4.4 Results

For rating agencies and traditional ABS investors, the *Probability of Default* (PD), that is the probability that the investors' principal will be reduced due to the occurrence of a catastrophic event, as well as the *Expected Loss* (EL), i.e. the expected percentage of the principal loss, are important comparative statistics. Furthermore, the corresponding spread level leading to an "actuarially fair" contract in the sense that the sum of the expected discounted cash flows equals zero is of interest, for example, in order to analyze risk premiums included in market prices.

In Table 4.8, the loss profiles as well as the respective spread levels are displayed for the two tranches within the Tartan deal and the three different historical parametrizations for the baseline component[32] P1, P2, and P3. The results clearly reflect the increased exposure to catastrophic mortality risk of the lower Class B

[32]While discussing the influence of the baseline component, we always assume the following parametrization for the catastrophe component: $\kappa = 4.6052$, $\lambda = 0.074$, $\zeta = 0.002799$ (cf. Table 4.5).

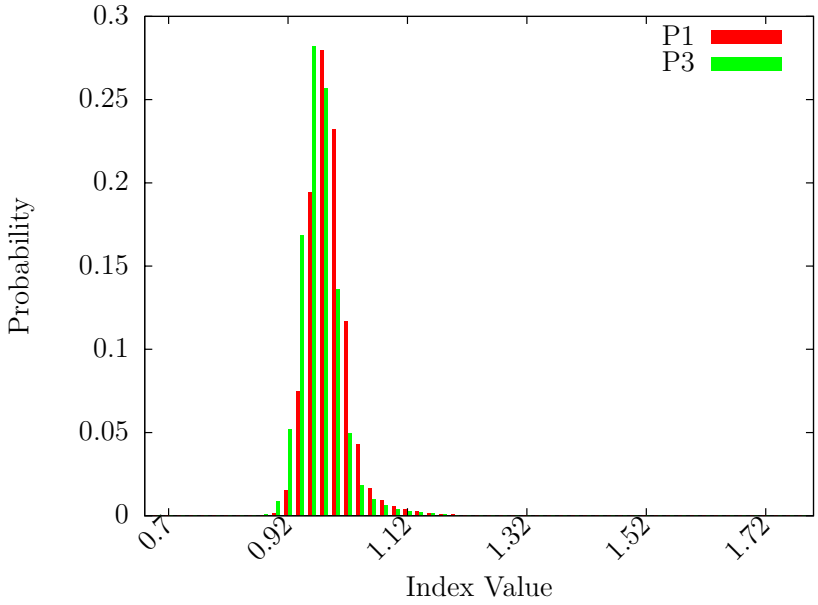


Figure 4.7: Estimated distribution of the index value for scenarios P1 and P3. P1 assumes only very mild mortality improvements in the future, while P3 presumes strong mortality improvements.

notes in comparison to the more senior Class A notes: All three risk measures are reduced by more than 63%. This does not seem surprising considering the fact that the lower tranche will be completely exhausted if the higher tranche is triggered. Moreover, all three quantities are relatively high for the more conservative parametrization P1, where only very mild future mortality improvements are assumed. While this general trend also does not seem peculiar, it occurs that the influence of the baseline component is quite pronounced: From parametrization P1 to P2, the expected loss is reduced by almost 16%, and from P2 to P3 the reduction is still about 10%. This reveals that despite the rather short maturity of three years the baseline component considerably affects the loss profile.

This does not mean that the bond may be triggered by an adverse evolution of the baseline component. In fact, when neglecting the catastrophe component, the default probability is at an almost negligible level for all three parametrizations of the baseline component – but differences in mortality improvements due to the parametrization of the baseline component affect the probability that the tranche is triggered given a "jump" occurred. In Figure 4.7, the discretized distribution of the index at time $t = 3$, i_3, is plotted for parametrizations P1 and P3. We find that the right tail is only slowly declining in comparison to the left tail. This "skewness" is due to the influence of the catastrophe component, i.e. the influence of the (positive) jumps of the mortality intensity. When comparing the distributions of the

index implied by the two baseline parametrizations, it appears that the shape is very similar but that more pronounced mortality improvements lead to a left shift. This means that for higher mortality improvements, catastrophic occurrences may be leveled to the point where they do not lead to a trigger event.

However, the catastrophe component is the more important risk driver for CATM securitizations. In particular, the uncertainties regarding the corresponding parameters have a significant effect: The loss profile of the bonds is very sensitive to changes in all three parameters affecting the catastrophe component.[33] As indicated in Section 4.3.1, there are particular problems when trying to find an adequate parametrization for the expected jump size ζ since different observation periods or different demographic structures may yield considerably different outcomes of the calibration procedure. In Table 4.9, comparative statistics for different parametrizations are presented. For the middle choice of the baseline component (P2), aside from risk measures resulting from parameters corresponding to the 2003 demographic structure (P2 and P2$'$) the results implied by the 1959 demographic structure (P21 and P21$'$) for exponentially as well as Gamma distributed jumps, respectively, are provided. Moreover, the loss profile for the Tartan bonds as quoted from Linfoot (2007) is shown.

We find that there are distinct differences between the two contemplated jump distributions, which emanate from a higher variance of the Gamma distributed jumps. For the utilized data consisting of only five model points, it is arguable whether the additional degree of freedom within the Gamma distribution is essential or even appropriate, but for larger data-sets it may prove necessary in order to obtain a significantly better fit to the empirical distribution. Moreover, in comparison to the jump size distribution, the influence of the population structure appears to be similarly pronounced. The expected loss is reduced by approximately 20% to 30% for the different tranches and jump distributions when considering the 2003 opposed to the 1959 demographic structure.

The risk measures from Linfoot (2007) are considerably lower than our results for the 2003 population, but when comparing the results for Class A and B notes as well as the ratios of default probabilities and expected losses within one set of results to our outcomes the implicit structure is quite similar. This observation indicates that the deviations of our findings in comparison to the "official" quotes from Linfoot (2007) do not result from a distinctive structural difference in the

[33] See Figures B.1, B.2, and B.3 in the Appendix for sensitivities of the expected tranche loss to changes in μ_0, κ, and ζ, respectively. The basic parametrization of the catastrophe component, which this sensitivity analysis is based on, is again $\kappa = 4.6052$, $\lambda = 0.074$, and $\zeta = 0.002799$ (cf. Table 4.5).

	Cl. B Tranche (110%-115%)			*Cl. A Tranche (115%-120%)*		
Scenario	PD(%)	EL(%)	Spread(bps)	PD(%)	EL (%)	Spread(bps)
P21	2.6464	1.7507	70.06	1.0854	0.7181	28.55
P21′	2.9376	2.1615	88.16	1.5460	1.1594	47.27
P2	2.0954	1.3223	52.53	0.7710	0.4856	19.20
P2′	2.4742	1.7639	71.74	1.2208	0.8841	35.87
Quoted	0.88	0.54	-	0.29	0.16	-

Table 4.9: Influence of the Catastrophe Component on the risk profile of the Tartan tranches. Calculated PDs, ELs and tranche spreads are based on the estimated parameters in the different scenarios. All scenarios presume medium mortality improvements in the future, but differ in assumed demographic structures and jump size distributions: P2 and P2′ rely on the 2003 demographic structure, while P21 and P21′ assume the 1959 demographic structure. P2 and P21 presume exponentially distributed jumps, P2′ and P21′ Gamma distributed jumps. Quoted values are from Linfoot (2007).

model specification but rather from differences in the considered parametrizations, in particular for the expected jump size ζ and the mean reversion parameter κ. Regarding the difficulties which come along with the calibration procedure, these deviations indicate that the estimates provided by risk modeling firms should be interpreted carefully by investors and, especially, rating agencies.

Even though the resulting spreads for our parametrizations are presumably higher than the spread levels corresponding to the loss profiles from Linfoot (2007), they are clearly still well below the market level of 300bps. Figure 4.8 shows the sensitivity of the spread level to changes in the expected jump size ζ. For relatively low values, the tranche spreads increase exponentially in the expected jump size.[34] However, as ζ increases the sensitivity lessens and the curve becomes, ceteris paribus, concave. This peculiarity is due to the structure of the deal: If a jump is large enough such that the complete principal is exhausted, it is not relevant by how much the jump exceeds this critical value. In particular, this means that the differences between the spread levels for the two different tranches decrease with an increasing expected jump size, as – beyond some critical point – most jumps will fully exhaust both tranches. Thus, we obtain an upper bound of approximately

[34]See also Figure B.3 in the Appendix for the same observation of the sensitivity of the expected tranche loss.

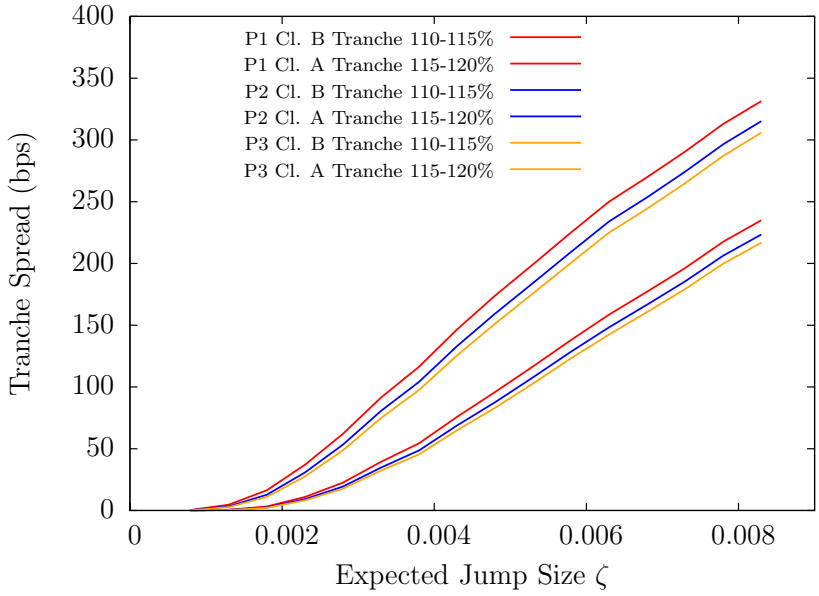


Figure 4.8: Influence of the expected jump size ζ on the Tartan tranche spreads.

1200bps for the spread level since for "infinitely" large jumps, the only question in view is whether a jump occurs or not which is controlled by the jump intensity $\mu_0 = 0.074$.

This structure is similar to the one of CDO tranches. In fact, the tranche loss distributions displayed in Figure 4.9 have exactly the same shape as for CDOs. The left tail of the index distribution is attributed to the 0% loss, whereas the outer right tail is attributed to a full loss. In particular, it is worth noting that the cumulated loss probabilities for the Class A tranche exceeding the 0% level add up to the full loss probability of the Class B tranche. Moreover, we can again observe the "shift" of the distributions when comparing the histograms for parametrization P1 and P3.

The relatively young history of the CATM market and the unfamiliarity of ABS investors with mortality contingent securities suggest that the spread levels investors can earn within CATM transactions are likely to be above the levels for CDOs with a similar PD and EL, i.e. with a similar rating (for a comparison see e.g. Logisch (2007)).[35] This idea is also backed by a comparison of the different CATM deals

[35] Although many market participants base their decisions merely on the rating, it is, of course, not sufficient to characterize the total risk of a security. For instance, senior CDO tranches are usually much more exposed to systemic risk than corporate bonds with the same rating suggesting higher spread levels of the tranches. Since CATM securities are low-beta investments, however, their high spread levels in comparison with traditional structured finance securities indicate that they include a novelty premium.

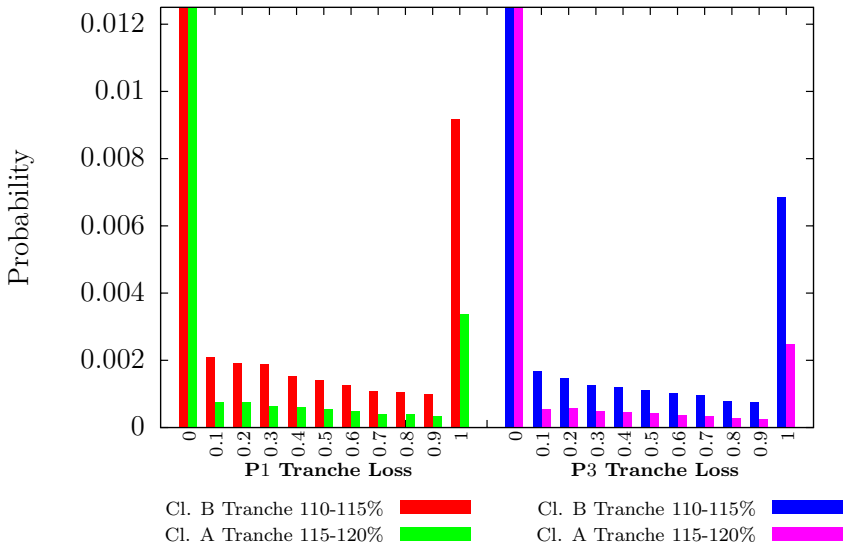


Figure 4.9: Discretized loss distributions of the Tartan tranches for scenarios $P1$ and $P3$. P1 assumes only very mild mortality improvements in the future, while P3 presumes strong mortality improvements. The probability of a 0% loss is capped at 0.0125.

so far. In Table 4.10, risk measures and spread levels as implied by the Tartan price (cf. Section 4.3.3) as well as the quoted market spreads for tranches from the Vita I and the Vita III transactions are displayed.[36] It appears that the quoted spread level within the Vita I deal was considerably higher than the spread level implied by the Tartan price. This does not seem surprising as Vita I was the first CATM transaction, and consequently the spreads included a considerable novelty premium. Conversely, the actual spread level for the B-II notes of the Vita III transaction is considerably lower than the respective spread implied by Tartan.

This relationship cannot be observed for the A-VII tranche: Here, the market spread slightly exceeds the Tartan-implied spread. However, the difference of the market spreads between the two Vita tranches seems rather small considering the significantly higher exposure of the Class B-II note to catastrophic occurrences. In particular, if we adjust the expected jump size to match one of the two prices, the other model-endogenous spread will be far from the observed one. However, we only adjust one parameter, namely the expected jump size; as depicted earlier in this section, the sensitivity of the spread level to the expected jump size fades

[36]Note that using the same parametrization would suggest that the underlying population is the same for all transactions, which is not the case. Thus, our findings have to be considered with care as e.g. diversification effects are not included.

Vita I

	Cl. A Tranche (130%-150%)		
	PD(%)	EL(%)	Spread(bps)
Calibration	2.7508	1.4408	47.12
Quoted	-	-	135

Vita III

	Cl. B-II Tranche (120%-125%)			*Cl. A-VII Tranche (125%-145%)*		
	PD(%)	EL(%)	Spread(bps)	PD(%)	EL(%)	Spread(bps)
Calibration	7.1754	4.9992	190.26	6.0290	2.6332	74.97
Quoted	-	-	112	-	-	80

Table 4.10: Summary of results for risk-adjusted parametrization based on the Tartan tranche. Quoted values are from Table 4.3.

for higher values of ζ (see Figure 4.8), whereas the spread level decreases exponentially in κ and increases almost linearly in the jump intensity μ_0.[37] Therefore, using our model, it is possible to mimic price structures as observed within the Vita III transaction by simultaneously adjusting several parameters instead of a single one. However, a calibration via Monte Carlo simulations as was carried out in Section 4.3.3 will become cumbersome. Thus, when calibrating our model to several tranche prices as e.g. within the Osiris or Vita III transaction the calibration procedure based on the approximative derivation of the index distribution from the Appendix, which is also adverted in Section 4.3.3, may be advisable.

Table 4.11 shows the results for the risk-adjusted parametrization based on insurance prices as was explained in Section 4.3.2. In comparison to the results from the real-world measure parametrizations, we find that all risk measures increase dramatically. Despite the problems with the calibration due to selection effects and differences in the populations considered as described in the foregoing section, these large deviations indicate that insurance prices include considerable margins for adverse mortality evolutions due to occurrences of catastrophic events. In particular, we find that the resulting spread level of approximately 277bps for the Class B notes is only slightly lower than the market spread of 300bps. Disregarding possible flaws, this means that Scottish Re was able to bin the catastrophic mortality risk from their books for less than 23bps since the 277bps constitute a

[37] See Figures B.1 and B.2 in the Appendix for the sensitivity of the expected tranche loss to changes in the jump intensity and the speed of mean reversion.

	Cl. B Tranche (110%-115%)			*Cl. A Tranche (115%-120%)*		
	PD(%)	EL(%)	Spread(bps)	PD(%)	EL (%)	Spread(bps)
Calibration	7.1272	6.6586	277.29	6.2092	5.8092	239.36
Quoted	-	-	300	-	-	-

Table 4.11: Summary of results for risk-adjusted parametrization based on insurance prices. Quoted values are from Table 4.3.

lower bound for the spread level implied by insurance prices (cf. 4.3.2). Furthermore, as detailed in Section 4.1.2 and indicated in Table 4.10, the Tartan deal was priced wider than many other deals, meaning that this difference may have even been smaller for other transactions. Hence, it is even conceivable that within some tighter priced notes, such as the Vita III Class B-II notes, (re)insurers were able to lay off their catastrophic mortality risk by earning rather than paying a premium.

These observations provide us with a possible answer to the question of why the CATM market has grown so quickly over the last years: For ABS investors, CATM transactions provide investment opportunities with a familiar payoff structure, but the spreads one can earn seem to exceed the ones within the credit market, possibly due to considerable novelty premiums. However, the margins for catastrophic mortality risk within insurance prices appear to be, if at all, only slightly lower than the margins within the CATM bonds. Thus, for (re)insurers CATM transactions seem to provide relatively cheap – or even profitable – means to remove catastrophic mortality risks from their liability side.

In this regard, it is worth noting that a growing CATM market with decreasing spread margins may eventually yield decreasing life insurance premiums, which in turn would induce welfare gains.

4.5 Summary

Catastrophe Mortality Bonds are a recent capital market innovation providing insurers and reinsurers with the possibility to transfer catastrophe mortality risk off their balance sheets to capital markets. While the various transactions differ in the composition of the underlying reference population, the basic structure is the same: Based on mortality data as reported by official entities, a combined mortality index is calculated. If this index exceeds a certain level, the bonds will be triggered and the investors' principal will be reduced. In return, investors receive

coupon payments on their principal including spread margins for the adopted risk, a basic structure which is similar to the one of CDO transactions.

So far, there have been five public deals in the market, four of which Swiss Re has been involved in as protection buyer and/or arranger. Since the first deal in late 2003, in which only one tranche with a relatively low risk exposure was issued, the market has developed considerably – more recent deals include several tranches with different seniorities. Moreover, the spectrum of investors has widened substantially: While the first bond was mainly sold within the insurance world or to specialized CAT bond investors, now several tranches are wrapped by monoline insurers and traditional ABS investors as well as hedge funds have found interest in these securities.

An important role in the arrangement and the execution of these transactions is played by so-called risk modeling firms, who are in charge of the calculation of the combined mortality index and the provision of comparative statistics such as default probabilities or expected losses of the securities for investors and rating agencies. Despite the growing market and the increasing bandwidth of investors, so far only one company, the actuarial consultant *Milliman*, was appointed as the calculation agent for all deals. However, within the last transaction (Vita III), the consulting firm *Risk Management Solutions* was hired as an adviser by one of the involved monoline insurers. Their modeling approaches differ considerably: While Milliman uses statistical forecasts based on an actuarial model, RMS relies on their expertise regarding pandemic occurrences in a causal modeling approach.

In this chapter, a time-continuous actuarial model for analyzing and pricing mortality contingent securities is introduced. The model consists of two additive parts: A baseline component, which models the "regular" random fluctuations of mortality over time and is driven by a diffusion, and a catastrophe component governed by a jump process. Due to its affine structure, survival probabilities can be determined analytically up to the solution of ordinary differential equations, and – on this basis – the "classical actuarial toolbox" can be used to determine insurance premiums, for example.

In order to apply this model for analyzing mortality contingent securities, it naturally needs to be calibrated. We provide a detailed discussion of different calibration procedures and resulting parametrizations. In addition to a calibration based on historical data, we derive risk-adjusted parametrizations based on insurance quotes and market prices of catastrophe mortality bonds, respectively.

Our discussion shows that finding adequate parameters based on the data used in practice is very difficult, particularly for the catastrophe component. Therefore, we

do not consider a single set of parameters but several parametrizations and conduct detailed sensitivity analyses. We find that the outcomes regarding expected losses and default probabilities of the considered securities differ significantly among the different sets of parameters, which leads to the conclusion that loss profiles as provided by the risk modeling firms have to be considered with care. In particular, the provided risk measures are substantially lower than our results for all considered parametrizations although there are no structural differences in the outcomes, which indicates that the parametrizations used by the calculation agents are rather "optimistic". A collaboration of actuaries and experts in epidemiological research may potentially lead to more reliable results.

Analyzing the loss distribution of the notes, one detects that they look very similar to loss distributions of CDO tranches, suggesting that ABS investors may feel quite comfortable with these securities as they are used to their structure. Moreover, when comparing loss probabilities and expected losses, the risk-return profile of CATM bonds seems to be very attractive. However, one has to keep in mind that the structure of the underlying risk is not alike meaning that comparisons based on low order moments or partial moments may be misleading. Nevertheless, a comparison of the pricing of the different transactions suggests that there is a substantial novelty premium included in the spread margins, which explains the investors' interest in the notes.

Comparing the spread margins to notional margins in term life insurance prices, the differences seem to be rather small. This indicates that insurers and reinsurers can take advantage of the risk transfer at a relatively low cost or even by earning a premium, which may explain the quick growth of the market from the insurer's perspective.

For assessing CATM bonds with more than one underlying population, as a next step, the model can be extended to multiple dimensions. As depicted in Section 4.3.3, the model structure and, in particular, the analytic properties will remain the same under certain assumptions on the dependence of the respective baseline components.

Chapter 5

Conclusion

This thesis introduces a mathematical framework for modeling a vector of stopping times, which serves as an abstract setup for modeling various real-world phenomena. In particular, several models that have been proposed in financial literature are included as special cases. After providing a detailed discussion of its properties, we apply this framework in order to model and empirically analyze two different classes of structured finance securities: Structured credit products and mortality contingent catastrophe bonds. A synopsis of the principal findings of these investigations can be found in Sections 2.9, 3.6 and 4.5.

We conclude by providing answers to the basic questions raised in the introduction of this thesis:

What types of models are needed to explain the characteristics of structured finance securities?

The clustering of defaults presents one of the major risk drivers for structured credit products. Nevertheless, our findings indicate that in order to explain this clustering, it is sufficient to model observable variables such as the firms Expected Default Frequency (EDF) or the S&P 500 index. More specifically, it does not appear necessary to include additional contagion effects or frailty variables.
When estimating different time-continuous default intensity models, we found that models with jumps are better capable of explaining the intensity dynamics. In particular, our analysis suggests including jumps when modeling firms of low creditworthiness.
From a practical perspective, however, more complex models do not necessarily lead to significantly different results than simple models. For instance, we show that simulated transition matrices of structured credit products relying on purely diffusion-based specifications and jump-diffusion models are quite similar.
Regarding the question of whether or not contagion effects should be incorporated to explain the risk characteristics of structured credit products, we observe that

in practical model applications, estimation errors are by far more influential than deviations stemming from the assumption of conditional independence. Even if contagion effects play a dominant role in the data-generating model, results derived with an estimated conditional independence model and results based on estimates of the true model may be very close.
For modeling mortality contingent catastrophe bonds, on the other hand, we show that an appropriate model should consist of at least two components: One relating to regular fluctuations of mortality rates and another one driven by jumps, where the latter represents catastrophic events. However, it is necessary to point out that the catastrophe component will always be subject to high parameter uncertainty since data on past events is sparse.

Do different models imply similar profiles for the securities? When do simple and complex models lead to comparable results?

Our theoretical investigation of model-implied dependence patterns reveals that many time-continuous models from scientific literature yield a similar dependence structure between the stopping times over a fixed time horizon, which can be described by a well-known copula class. Moreover, by simulating transition matrices of structured credit products, we found that estimated models with and without jumps yield similar results. We also demonstrate that the issue of contagion effects is of lower relevance in comparison with general problems applying to all models such as estimation errors. A suitable conditional independence model will usually lead to risk profiles that are close to those obtained from an estimated contagion model as long as the distributional properties of the integrated portfolio intensity are described sufficiently well.

However, while these observations support the continued use of standard models, our results do not imply that models with contagion effects or intensity jumps are generally irrelevant when modeling structured credit products. It is conceivable that such effects only appear small in the data set that we examined, which comprised rated corporate bond issuers with traded equity. For example, a bank wishing to assess the risk of a structured credit security referencing a private loan portfolio or a portfolio of residential mortgages may be well advised to take intensity jumps into account. Future research should examine whether the results of this thesis carry over to such more general situations.

Appendix A

Appendix to Chapter 3

The BAJD model: Closed-form Solutions of an Important Transform and an Exact Simulation Algorithm

This part of the appendix states a closed-form expression for an important transform of a process λ which evolves according to the SDE (2.27), i.e. follows a BAJD model. Moreover, we derive an exact simulation algorithm for λ. This algorithm extends the exact simulation algorithm for the CIR model that can be found in Glasserman (2004), p. 124.

Let us first consider the calculation of the following transform for which we obtain based on Theorem 2.7.1

$$ {}_{T,t}\varphi_{\int_0^\bullet \lambda(s)ds,\lambda(\bullet)}(c_1',c_2') := \mathbb{E}\left[e^{-c_1'\int_t^T \lambda(s)ds - c_2'(\lambda(T)-\lambda(t))} \middle| \lambda(t) \right] = e^{u'(t,T)+(v'(t,T)+c_2')\lambda(t)} $$

where (c_1', c_2') as in Theorem 2.7.1 and u' and v' are solutions of the following ODE system:

$$ \begin{aligned} \dot{u}' &= -v'\kappa\eta - \mu\frac{\zeta v'}{1-\zeta v'}, \\ \dot{v}' &= c_1' + v'\kappa - 0.5(v')^2\sigma^2 \end{aligned} $$

with terminal conditions $u'(T,T)=0$, $v'(T,T)=-c_2'$. By substituting $v(T-t) := v'(t,T)$ and $u(T-t) := u'(t,T)$ and setting $c_1 := -c_1'$ and $c_2 := -c_2'$, we finally obtain

$$ \begin{aligned} \dot{u} &= v\kappa\eta + \mu\frac{\zeta v}{1-\zeta v}, \\ \dot{v} &= c_1 - v\kappa + 0.5v^2\sigma^2 \end{aligned} $$

with initial conditions $u(0) = 0$ and $v(0) = c_2$. Duffie and Gârleanu (2001) provide the following explicit solutions for this ODE system:

$$\begin{aligned} u(T-t) &= \frac{\eta\kappa(a_1e_1 - d_1)}{\gamma e_1 d_1} \log \frac{e_1 + d_1 e^{\gamma(T-t)}}{e_1 + d_1} + \frac{\eta\kappa}{e_1}(T-t) \\ &\quad + \frac{\mu(a_2e_2 - d_2)}{\gamma e_2 d_2} \log \frac{e_2 + d_2 e^{\gamma(T-t)}}{e_2 + d_2} + \left(\frac{\mu}{e_2} - \mu\right)(T-t) \\ v(T-t) &= \frac{1 + a_1 e^{\gamma(T-t)}}{e_1 + d_1 e^{\gamma(T-t)}} \end{aligned}$$

with

$$\begin{aligned} e_1 &= \frac{\kappa + \sqrt{\kappa^2 - 2\sigma^2 c_1}}{2c_1} & e_2 &= 1 - \frac{\zeta}{e_1} \\ d_1 &= (1 - e_1 c_2)\frac{-\kappa + \sigma^2 c_2 + \sqrt{\kappa^2 - 2\sigma^2 c_1}}{-2\kappa c_2 + \sigma^2 c_2^2 + 2c_1} & a_1 &= (d_1 + e_1)c_2 - 1 \\ \gamma &= \frac{d_1(-\kappa + 2c_1 e_1) + a_1(-\kappa e_1 + \sigma^2)}{a_1 e_1 - d_1} & d_2 &= \frac{d_1 - \zeta a_1}{e_1} \\ a_2 &= \frac{d_1}{e_1} \end{aligned}$$

Therefore,

$${}_{T,t}\varphi_{\int_0^\bullet \lambda(s)ds, \lambda(\bullet)}(c_1', c_2') = e^{u(T-t) + (v(T-t) - c_2)\lambda(t)}$$

Subsequently, we present an exact simulation algorithm for a process λ following a BAJD model.

Algorithm A.0.1 *Assume that λ evolves according to the SDE (2.27) and let $\nu := \frac{4\eta\kappa}{\sigma^2}$. Then, in order to draw a path of λ on the time grid $0 = t_0 < t_1 < \ldots < t_K$ proceed as follows*

CASE $\nu > 1$*:*
  Specify $\lambda(0)$
  for $i = 0, \ldots, K-1$
    generate $M \sim Poi\left(\mu\left(t_{i+1} - t_i\right)\right)$
    $U_0 \leftarrow 0$
    for $j = 1, \ldots, M$
      generate $U_j \sim U_{[0,(t_{i+1} - t_i)]}$
    end
    $\lambda_0^* \leftarrow \lambda(t_i)$
    for $j = 1, \ldots, M$
      $c \leftarrow \sigma^2 \left(1 - e^{-\kappa\left(U_{(i)} - U_{(i-1)}\right)}\right)/(4\kappa)$
      $p \leftarrow \lambda_{j-1}^* \left(e^{-\kappa\left(U_{(i)} - U_{(i-1)}\right)}\right)/c$
      generate $Z \sim N_{0,1}$

$$
\begin{aligned}
&\qquad\qquad \textit{generate } X \sim \chi^2_{\nu-1} \\
&\qquad\qquad \textit{generate } Y \sim Exp\left(\tfrac{1}{\zeta}\right) \\
&\qquad\qquad \lambda^*_j \leftarrow \left((Z+\sqrt{p})^2 + X + Y\right) \\
&\qquad \textit{end} \\
&\qquad c \leftarrow \sigma^2\left(1 - e^{-\kappa(t_{i+1}-U_{(M)})}\right)/(4\kappa) \\
&\qquad p \leftarrow \lambda^*_M\left(e^{-\kappa(t_{i+1}-U_{(M)})}\right)/c \\
&\qquad \textit{generate } Z \sim N_{0,1} \\
&\qquad \textit{generate } X \sim \chi^2_{\nu-1} \\
&\qquad \lambda(t_{i+1}) \leftarrow \left((Z+\sqrt{p})^2 + X\right) \\
&\textit{end}
\end{aligned}
$$

$$
\begin{aligned}
&\textit{CASE} \quad 0 < \nu \leq 1\textit{:} \\
&\quad \textit{Specify } \lambda(0) \\
&\quad \textit{for} \quad i = 0, \ldots, K-1 \\
&\qquad \textit{generate } M \sim Poi\left(\mu\left(t_{i+1} - t_i\right)\right) \\
&\qquad U_0 \leftarrow 0 \\
&\qquad \textit{for} \quad j = 1, \ldots, M \\
&\qquad\qquad \textit{generate } U_j \sim U_{[0,(t_{i+1}-t_i)]} \\
&\qquad \textit{end} \\
&\qquad \lambda^*_0 \leftarrow \lambda(t_i) \\
&\qquad \textit{for} \quad j = 1, \ldots, M \\
&\qquad\qquad c \leftarrow \sigma^2\left(1 - e^{-\kappa\left(U_{(i)}-U_{(i-1)}\right)}\right)/(4\kappa) \\
&\qquad\qquad p \leftarrow \lambda^*_{j-1}\left(e^{-\kappa\left(U_{(i)}-U_{(i-1)}\right)}\right)/c \\
&\qquad\qquad \textit{generate } N \sim Poi\left(\tfrac{p}{2}\right) \\
&\qquad\qquad \textit{generate } X \sim \chi^2_{\nu+2N} \\
&\qquad\qquad \textit{generate } Y \sim Exp\left(\tfrac{1}{\zeta}\right) \\
&\qquad\qquad \lambda^*_j \leftarrow (cX + Y) \\
&\qquad \textit{end} \\
&\qquad c \leftarrow \sigma^2\left(1 - e^{-\kappa\left(t_{i+1}-U_{(M)}\right)}\right)/(4\kappa) \\
&\qquad p \leftarrow \lambda^*_M\left(e^{-\kappa\left(t_{i+1}-U_{(M)}\right)}\right)/c \\
&\qquad \textit{generate } N \sim Poi\left(\tfrac{p}{2}\right) \\
&\qquad \textit{generate } X \sim \chi^2_{\nu+2N} \\
&\qquad \lambda(t_{i+1}) \leftarrow cX \\
&\textit{end}
\end{aligned}
$$

The developed simulation Algorithm A.0.1 extends the exact simulation algorithm for a CIR process that can be found in Glasserman (2004), p. 124. Our algorithm exploits that, in case of a BAJD model, jumps occur independently of the Brownian component and that between jump arrivals the increments of the process are distributed as the increments of a CIR process. Furthermore, it is a well-known

fact that the increments of a CIR process stem from a non-central χ^2-distribution with ν degrees of freedom. Depending on the value of ν, one either has to combine a standard-normally and a χ^2-distributed, or a Poisson- and a χ^2-distributed random variable in order to simulate a non-central χ^2-distribution (for details see Glasserman (2004)).

A Simulation Algorithm for the Contagion Model given by the SDEs (3.22) and (3.23)

Exact simulation of processes λ and L that evolve according to the SDEs (3.22) and (3.23) is not feasible. In this case, we therefore rely on an Euler discretization of the SDEs.

Algorithm A.0.2 *Assume that λ and L evolve according to the SDEs (3.22) and (3.23). Then, in order to draw a path of λ and the loss process $L(t) = \frac{1}{I}\sum_{i=1}^{I} N_i(t)$ on the time grid $0 = t_0 < t_1 < \ldots < t_K$ proceed as follows:*

$L(0) \leftarrow 0$
Specify $\lambda(0)$
for $i = 0, \ldots, K-1$
generate $M \sim CIR\left(\lambda(t_i), \eta, \kappa, \sigma, t_{i+1} - t_i\right)$ *(based on Algorithm A.0.1)*
generate $Z \sim BERNOULLI\left(1 - \exp\left(-I(1 - L(t_i))\lambda(t_i)\left(t_{i+1} - t_i\right)\right)\right)$
generate $Y \sim Exp\left(\frac{1}{\zeta}\right)$
$\lambda(t_{i+1}) \leftarrow M + Y \cdot Z$
$L(t_{i+1}) \leftarrow L(t_i) + \frac{1}{I}Z$
end

"CIR" in the algorithm denotes that we draw – based on the exact simulation Algorithm A.0.1 – a single realization of a CIR process at t_{i+1}, given an initial value of $\lambda(t_i)$. Therefore, if no jumps are observed the discretization error implied by our algorithm will be zero, because errors are generally related to the jump part: First, in our algorithm a jump within $[t_i, t_{i+1}]$ occurs independently from the continuous part of the process. Second the number of jumps in each time interval is bounded by one. However, discretization errors are minor as long as fine enough time grids are considered.

A Simulation Algorithm for the SAJDM Model

Exact simulation of a process λ that evolves according to the SDE (3.11) is not feasible. In this case, we therefore rely on an Euler discretization of the SDE.

Algorithm A.0.3 *Assume that λ evolves according to the SDE (3.11). Then, in order to draw a path of λ on the time grid $0 = t_0 < t_1 < \ldots < t_K$ proceed as follows:*

Specify $\lambda(0)$
for $i = 0, \ldots, K-1$
generate $M \sim CIR(\lambda(t_i), \eta, \kappa, \sigma, t_{i+1} - t_i)$ *(based on Algorithm A.0.1)*
generate $Z \sim BERNOULLI\left(1 - \exp\left(-\left(\mu + \xi^{(1)}\lambda(t_i) + \xi^{(2)}H(t_i)\right)(t_{i+1} - t_i)\right)\right)$
generate $Y \sim Exp\left(\frac{1}{\zeta}\right)$
$\lambda(t_{i+1}) \leftarrow M + Y \cdot Z$
$H(t_{i+1}) \leftarrow \frac{1}{\epsilon}\lambda(t_0)e^{-t_{i+1}} + \int_0^{t_{i+1}} e^{-\epsilon(t_{i+1}-s)}\lambda(s)ds$
end

Like in Algorithm A.0.2, "CIR" denotes that we draw – based on the exact simulation Algorithm A.0.1 – a single realization of a CIR process at t_{i+1}, given an initial value of $\lambda(t_i)$. Therefore, if no jumps are observed the discretization error implied by our algorithm will be zero, because errors are generally related to the jump part: First, in our algorithm a jump within $[t_i, t_{i+1}]$ occurs independently from the continuous part of the process. Second the number of jumps in each time interval is bounded by one. However, discretization errors are minor as long as fine enough time grids are considered.

Implementation of the EM algorithm of Section 3.5.1

In our implementation of the EM algorithm, the number of simulations increases quadratically in each step as suggested by Cappé et al. (2005). We consider a maximum simulation number of 7500. As usual, to solve the problem of local maximums we run the algorithm a couple of times – each time with a different starting value.

The simulation of the conditional density in our second estimation approach is straight-forward and has already been described in the main text. In the following, we show how the regime process can be simulated that governs the factor in the first estimation approach. For ease of exposition, we demonstrate this for the BAJD model.

Let z_k and λ_k^c denote the realization of the regime process and the factor at t_k with $k \in \{1, \ldots, K\}$. In addition, we set $\Delta = t_{k+1} - t_k = t_k - t_{k-1}$ for all $k \in \{2, \ldots, K-1\}$. To simulate a whole path of the unobserved regime process, we apply the Gibbs sampler, which breaks the difficult task of simulating $z = (z_1, \ldots, z_K)$ given $\lambda^c = (\lambda_1^c, \ldots, \lambda_K^c)$ down into simulating the z_ks given $(z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_K)$ and $\lambda^c = (\lambda_1^c, \ldots, \lambda_K^c)$ (see Robert and Casella (1999), pp. 286).

Latter simulation is feasible because according to Bayes theorem we have that[1]

$$P\left(Z_k|z_{k-1},z_{k+1},\lambda^c_{k-1},\lambda^c_k,\lambda^c_{k+1}\right) \propto \prod_{l=k-1,k}\Bigg(\left(1-e^{-\Delta z_l \varpi\mu}\right)h(o_l^{(1)},o_l^{(2)}) \\ +e^{-\Delta z_l\varpi\mu} f_{N_{\kappa(z_l\varpi\eta-\lambda^c_l)\Delta,\sigma\sqrt{\lambda^c_l\Delta}}}\left(\lambda^c_{l+1}-\lambda^c_l\right)\Bigg) \\ \cdot \prod_{l=k-1,k}\left(\left(1-e^{-\Delta\mathcal{Q}_{12}}\right)1_{z_l\neq z_{l+1}}+e^{-\Delta\mathcal{Q}_{12}}1_{z_l=z_{l+1}}\right)$$

where

$$\begin{aligned} o_l^{(1)} : &= \lambda^c_{l+1}-\lambda^c_l-\kappa(\eta-\lambda^c_l)\Delta \\ o_l^{(2)} : &= \sigma\sqrt{\lambda^c_l\Delta} \end{aligned}$$

with appropriate modifications for the terminal cases Z_1 and Z_K.

Calculation of the tranches' expected loss in Section 3.5.3

In order to calculate the portfolio loss distribution, we apply the method described in Subsection 2.7.2. Since we assume homogeneity of the objects and work with a conditional independence model, the portfolio loss conditional on the integrated factor $\int_0^t \lambda^c(s)ds$ follows a binomial distribution. The density of the integrated factor is derived by Fourier inversion (see equation (2.24)) based on Fast Fourier Transforms (FFTs).[2] In the most general case, we proceed as follows:

1. We evaluate the characteristic function of the integrated factor on an unequally spaced grid of length 2^{16} by solving the ODE system stated in Proposition 3.2.1. We use a grid that is equally-spaced on a logarithmic scale. The mesh size of this equally-spaced grid depends on the variance of the integrated factor. For dispersed integrated factors, for example, we use a smaller mesh size.

2. The computed values of the characteristic function are then interpolated on an equally-spaced grid of length 2^{18} using a cubic-spline.

3. By applying the FFT, we finally get the density of the integrated factor.

Using the described interpolation is inspired by Eckner (2007). It is only applied in the models in which no analytic solution of the characteristic function is available,

[1] In our model implementation we presumed that $\mathcal{Q}_{12}=\mathcal{Q}_{21}$ in order to reduce the dimension of the parameter space for the optimization. The $h(\cdot,\cdot)$ function is defined in equation (3.19) on p. 111.

[2] We used the FFT implementation of the GSL.

i.e. only for the SAJD model (note that the other models without analytical solutions were rejected by a standard significance test). Otherwise, the characteristic function is directly evaluated on an equally-spaced grid of length 2^{18}. Apart from the general reduction of evaluation points, another effect of using an interpolation is that the number of *large* points for which the characteristic function has be evaluated is diminished, too.[3] The solution of the ODE system for these large points is computationally burdensome. Since the interpolation introduces an error, the number of points at which the characteristic function is originally evaluated should not be too small. Our choice of 2^{16} points guarantees that the computation of the density is not subject to significant interpolation errors.

[3]Note that due to the applied exponential transform, the mesh size between the evaluation points is closest at 0

Appendix B

Appendix to Chapter 4

Approximative Derivation of the Joint Distribution of i_2 and i_3, and Additional Figures

By equation (4.13), we have

$$\begin{aligned}
i_t &= \frac{1}{i_0} - \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} \left(e^{-\int_0^1 \lambda(x+s,t-1+s)\,ds} + e^{-\int_0^1 \lambda(x+s,t+s)\,ds} \right) \\
&\overset{\text{Taylor}}{\approx} \frac{1}{i_0} - \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} \left(\left(1 - \int_0^1 \lambda(x+s,t-1+s)\,ds \right) + \left(1 - \int_0^1 \lambda(x+s,t+s)\,ds \right) \right) \\
&= \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} \int_0^1 Y_2(t-1+s)\,ds + \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} \int_0^1 Y_2(t+s)\,ds \\
&\quad + \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} \int_0^1 Y_1(t-1+s) e^{b(x+s)+c}\,ds + \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} \int_0^1 Y_1(t+s) e^{b(x+s)+c}\,ds \\
&= \frac{1}{2i_0} \left(\int_{t-1}^t Y_2(s)\,ds + \int_t^{t+1} Y_2(s)\,ds \right) + \int_{t-1}^t Y_1(s) e^{bs}\,ds \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} e^{b(x-t+1)+c} \\
&\quad + \int_t^{t+1} Y_1(s) e^{bs}\,ds \sum_{\text{all}x} \frac{\omega_{x,m}}{2i_0} e^{b(x-t)+c},
\end{aligned}$$

where we used the simple Taylor expansion $e^x = (1+x)$ for x close to one. As the relevant quantities, in particular for the most relevant ages, are very small, the approximation will be close. In order to compute the (joint) distribution of i_2 and i_3, we need to compute the joint distributions of the random variables

$$\begin{aligned}
\text{Continuous Part} &: \Xi_1 := \int_0^1 Y_1(s) e^{bs}\,ds,\ \Xi_2 := \int_1^2 Y_1(s) e^{bs}\,ds,\ \Xi_3 := \int_2^3 Y_1(s) e^{bs}\,ds \\
\text{Jump Part} &: \Theta_1 = \int_0^1 Y_2(s)\,ds,\ \Theta_2 = \int_1^2 Y_2(s)\,ds,\ \Theta_3 = \int_2^3 Y_2(s)\,ds
\end{aligned}$$

with $Y_1(0) > 0$ and $Y_2(0) = 0$. As the jump part and the continuous part are independent, it is sufficient to derive the joint densities for Ξ_1, Ξ_2, and Ξ_3 as well as for Θ_1, Θ_2, and Θ_3, respectively. The joint density will be given by the product of the two densities. We will focus on the continuous part Y_1, but the jump part may be considered analogously.

It is important to note that Ξ_1, Ξ_2, and Ξ_3 are independent given $Y_1(1)$ and $Y_1(2)$. Therefore, for Borel sets $B_1,\ B_2,\ B_3,\ \tilde{B}_1,\ \tilde{B}_2$ we have

$$\begin{aligned}
& \mathbb{P}\left(\Xi_1 \in B_1, Y_1(1) \in \tilde{B}_1, \Xi_2 \in B_2, Y_1(2) \in \tilde{B}_2, \Xi_3 \in B_3\right) \\
= \ & \mathbb{P}\left(\Xi_1 \in B_1, \Xi_2 \in B_2, \Xi_3 \in B_3 \middle| Y_1(1) \in \tilde{B}_1, Y_1(2) \in \tilde{B}_2\right) \mathbb{P}\left(Y_1(1) \in \tilde{B}_1, Y_1(2) \in \tilde{B}_2\right) \\
= \ & \mathbb{P}\left(\Xi_1 \in B_1 \middle| Y_1(1) \in \tilde{B}_1, Y_1(2) \in \tilde{B}_2\right) \mathbb{P}\left(\Xi_2 \in B_2 \middle| Y_1(1) \in \tilde{B}_1, Y_1(2) \in \tilde{B}_2\right) \\
& \mathbb{P}\left(\Xi_3 \in B_3 \middle| Y_1(1) \in \tilde{B}_1, Y_1(2) \in \tilde{B}_2\right) \mathbb{P}\left(Y_1(1) \in \tilde{B}_1, Y_1(2) \in \tilde{B}_2\right) \\
= \ & \mathbb{P}\left(\Xi_1 \in B_1, Y_1(1) \in \tilde{B}_1\right) \mathbb{P}\left(\Xi_2 \in B_2, Y_1(2) \in \tilde{B}_2 \middle| Y_1(1) \in \tilde{B}_1\right) \mathbb{P}\left(\Xi_3 \in B_3 \middle| Y_1(2) \in \tilde{B}_2\right).
\end{aligned}$$

So, the joint density $f_{\Xi_1,\Xi_2,\Xi_3}(x,y,z)$ can be derived from the (conditional) densities $f_{\Xi_1,Y_1(1)}(x,y)$, $f_{\Xi_2,Y_1(2)|Y_1(1)}(x,y \mid z)$, and $f_{\Xi_3|Y_1(2)}(x \mid y)$ by

$$f_{\Xi_1,\Xi_2,\Xi_3}(x,y,z) = \int_{\mathbb{R}} \int_{\mathbb{R}} f_{\Xi_1,Y_1(1)}(x,u) f_{\Xi_2,Y_1(2)|Y_1(1)}(y,v \mid u) f_{\Xi_3|Y_1(2)}(z \mid v)\, du dv.$$

Since we are working in an affine framework, we can derive the (joint) Laplace or Fourier transforms for $\Xi_1, Y_1(1)$, for $\Xi_2, Y_1(2)$ given $Y_1(1)$, and for Ξ_3 given $Y_1(2)$ analytically up to the solution of ODEs similarly to the derivation of equation 4.7 above (see Duffie et al. (2000) for details). From these, we can derive the corresponding densities by inverting the respective transform (see e.g. Petrella (2004)) and compute the joint density of $f_{\Xi_1,\Xi_2,\Xi_3}(x,y,z)$ as above; thus, we are given the (joint) distribution of $i_2,\ i_3$ as deterministic functions of Ξ_1, Ξ_2, and Ξ_3.

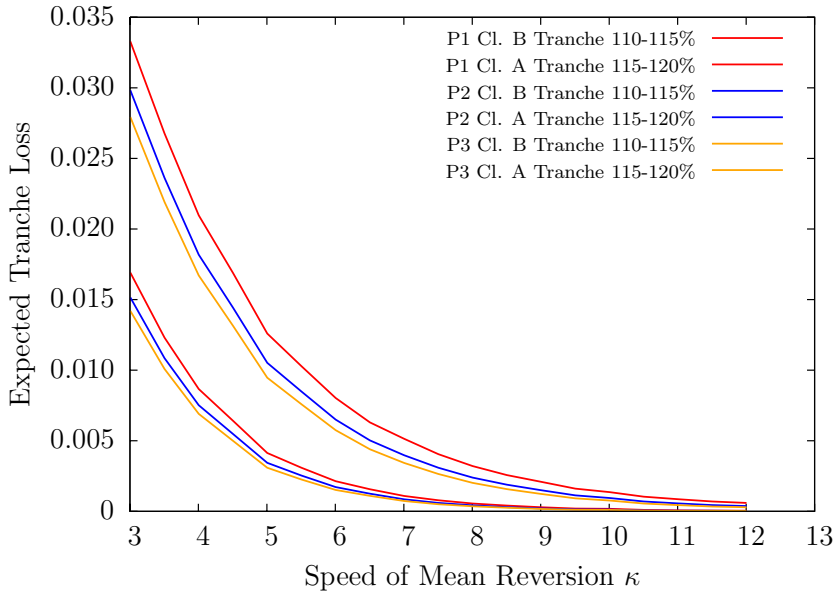


Figure B.1: Influence of the speed of mean reversion κ on the expected tranche loss.

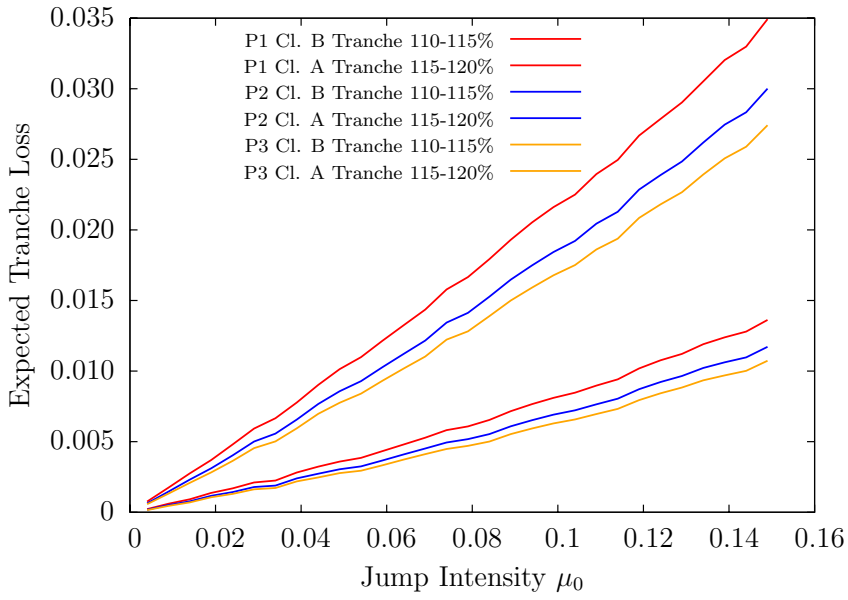


Figure B.2: Influence of the jump intensity μ_0 on the expected tranche loss.

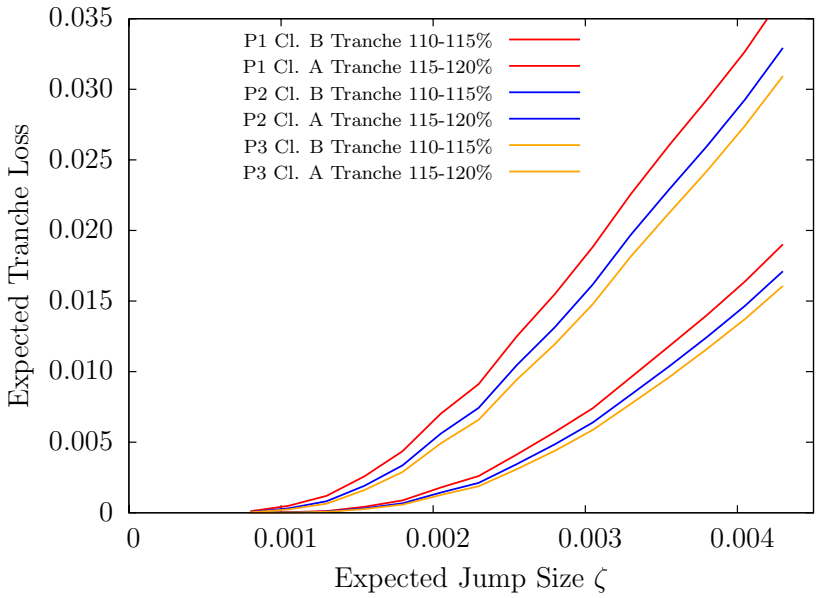


Figure B.3: Influence of the expected jump size ζ on the expected tranche loss.

Bibliography

American Academy of Actuaries, 2002. Final report of the american academy of actuaries commissioners standard ordinary task force. Presented in draft to the National Association of Insurance Commissioners Life and Health Actuarial Task Force, Philadelphia, PA.

Andersen, L., Sidenius, J., Basu, S., 2003. All your hedges in one basket. Risk, November 2003: 67–72.

Artzner, P., Delbaen, F., 1995. Default risk insurance and incomplete markets. Mathematical Finance, 5: 187–195.

Aven, T., 1985. A theorem for determining the compensator of a counting process. Scandinavian Journal of Statistics, 12: 69–72.

Bansal, R., Zhou, H., 2002. Term structure of interest rates with regime shifts. The Journal of Finance, 57: 1997–2043.

Bantwal, V.J., Kunreuther, H.C., 1999. A cat bond premium puzzle?. The Journal of Psychology and Financial Markets, 1: 76–91.

Bauer, D., Russ, J., 2006. Pricing longevity bonds using implied survival probabilities. Working Paper. Ulm University and Institut für Finanz- und Aktuarwissenschaften. Available at: www.mortalityrisk.org.

Benati, L., Goodhart, C., 2007. Investigating time-variation in the marginal predictive power of the yield spread. Working Paper. ECB.

Berndt, A., Douglas, R., Duffie, D., Ferguson, M., Schranz, D., 2005. Measuring default risk premia from default swap rates and EDFs. Working Paper. Available at: www.defaultrisk.com.

Bielecki, T.R., Rutkowski, M., 2002. Credit risk: Modeling, valuation and hedging. Springer Finance. Springer, London.

Biffis, E., 2005. Affine processes for dynamic mortality and actuarial valuation. Insurance: Mathematics and Economics, 37: 443–468.

Biffis, E., Denuit, M., Devolder, P., 2005. Stochastic mortality under measure change. Working Paper. CASS Business School and Université Catholique de Louvain.

Bingham, N., Kiesel, R., 2003. Risk-neutral valuation. Springer Finance. Springer, London.

Björk, T., 1999. Arbitrage theory in continuous time. Oxford University Press, Oxford, UK.

Black, F., Cox, J.C., 1976. Valuing corporate securities: Some effects of bond indenture provisions. The Journal of Finance, 31: 351–367.

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., Nesbitt, C.J., 1997. Actuarial mathematics. The Society of Actuaries, Schaumburg, Il, USA.

Brémaud, P., 1981. Point processes and queues. Springer Series in Statistics. Springer, New York.

Brown, T.C., Nair, M.G., 1988. A simple proof of the multivariate random time change theorem for point processes. Journal of Applied Probability, 25: 210–214.

Burtschell, X., Gregory, J., Laurent, J.-P., 2007. A comparative analysis of CDO pricing models. Working Paper. Available at: www.defaultrisk.com.

Cairns, A.J., Blake, D., Dowd, K., 2006. Pricing death: Frameworks for the valuation and securitization of mortality risk. ASTIN Bulletin, 36: 79–120.

Cappé, O., Moulines E., Rydén, T., 2005. Inference in hidden Markov models. Springer Series in Statistics. Springer, New York.

Chapovsky, A., Rennie, A., Tavares, P., 2006. Stochastic intensity modelling for structured credit exotics. Working Paper. Available at: www.defaultrisk.com.

Carr, P., Madan, D.B., 1999. Option valuation using the fast Fourier transform. Journal of Computational Finance, 2: 61–73.

Chen, L., Filipović, D., 2007. Credit derivatives in an affine framework. Asia-Pacific Financial Markets, 14: 123–140.

Cochran, W.G., 1954. Some methods of strengthening χ^2 tests. Biometrics, 10: 417–451.

Collin-Dufresne, P., Goldstein, R., Hugonnier, J., 2004. A general formula for valuing defaultable securities. Econometrica, 72: 1377–1407.

Cont, R., Tankov, P., 2004. Financial modelling with jump processes. Chapman & Hall/CRC, Boca Raton.

Cousin, A., Laurent, J.-P., 2007. Comparison results for credit risk portfolios. Forthcoming in Insurance: Mathematics and Economics.

Cowley, A., Cummins, J.D., 2005. Securitization of life insurance assets and liabilities. The Journal of Risk and Insurance, 72: 193–226.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica, 53: 385–407.

Cox, S.H., Lin, Y., Wang, S., 2006. Multivariate exponential tilting and pricing implications for mortality securitization. The Journal of Risk and Insurance, 73: 719–736.

Cox, S.H., Pedersen, H.W., Fairchild, J.R., 2000. Economic aspects of securitization of risk. ASTIN Bulletin, 30: 157–193.

Crouhy, M., Galai, D., Mark, R., 2000. A comparative analysis of current credit risk models. Journal of Banking and Finance, 24: 59–117.

Cummins, J.D., 2006. Should the government provide insurance for catastrophes?. Federal Reserve Bank of St. Louis Review, 88: 337–379.

Culp, C.L., 2006. Structured finance and insurance. Wiley, Hoboken.

Dahl, M., 2004. Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts. Insurance: Mathematics and Economics, 35: 113–136.

Dahl, M., Melchior, M., Møller, T., 2006. On systematic mortality risk and risk-minimization with survivor swaps. Working Paper. Nordea Markets and PFA Pension, Copenhagen. Available at: www.mortalityrisk.org.

Das, S., Duffie, D., Kapadia, N., Saita, L., 2007. Common failings: How corporate defaults are correlated. The Journal of Finance, 62: 93–117.

Deacon, J., 2004. Global securitisation and CDOs. Wiley, Hoboken, N.J..

Dellacherie, C., Meyer, P.-A., 1980. Probabilités et Potentiels. Chapitres 5 à 8: Théorie des Martingales. Actualités Scientifiques et Industrielles. Vol. 1385. Hermann, Paris.

Dempster, A., Laird, N., Rubin, D., 1985. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society Series B, 39: 1–38.

Doherty, N., 1997. Financial innovations in the management of catastrophic risk. Journal of Applied Corporate Finance, 10: 84–95.

Duffie, D., Gârleanu N., 2001. Risk and valuation of collateralized debt obligations. Financial Analysts Journal, 57: 41–59.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica, 68: 1343–1376.

Duffie, D., Lando, D., 2001. Term structures of credit spreads with incomplete accounting information. Econometrica, 69: 633–664.

Duffie, D., Filipović, D., Schachermayer, W., 2003. Affine processes and applications to finance. The Annals of Applied Probability, 13: 984–1053.

Duffie, D., Saita, L., Wang, K., 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics, 83: 635–665.

Duffie, D., 2007. Innovations in credit risk transfer: Implications for financial stability. Working Paper. Available at: www.defaultrisk.com.

Eckner, A., 2007. Computational techniques for basic affine models of portfolio credit risk. Working Paper. Available at: www.defaultrisk.com.

Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: Properties and pitfalls. In: "Risk Management: Value at Risk and Beyond". Dempster, M.A.H., Moffatt, H.K., (Eds.). Cambridge University Press.

Equitable, 2004. Equitable accumulator product prospectus. The Equitable Life Assurance Society of the United States, New York.

Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. The Journal of Finance, 58: 1269–1300.

Euroweek, 2006. Capital markets shield AXA from extreme mortality risk. Euroweek, 978(11/2006): 62.

Feldhütter, P., 2008. An empirical investigation of an intensity-based model for pricing CDO tranches. Working Paper. Available at: www.defaultrisk.com.

Fender, I., Mitchell, J., 2005. Structured finance: complexity, risk and the use of ratings. BIS Quarterly Review, June 2005: 67–79.

Frey, R., Backhaus, J., 2007. Pricing and hedging of portfolio credit derivatives with interacting default intensities. Forthcoming in International Journal of Theoretical and Applied Finance.

Froot, K.A., 2001. The market for catastrophic risk: A clinical examination. Journal of Financial Economics, 60: 529–571.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6: 721–741.

Georges, P., Lamy, A.-G., Nicolas, E., Quibel, G., Roncalli, T., 2001. Multivariate survival modelling: A unified approach with copulas. Working Paper. Available at SSRN: http://ssrn.com/abstract=1032559.

Glasserman, P., 2004. Monte carlo methods in financial engineering. Springer, New York.

Gompertz, B., 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London, 115: 513–585.

Gordy, M., 2003. A risk-factor model foundation for ratings-based bank capital rules. Journal of Financial Intermediation, 12: 199–232.

Graziano, G.D., Rogers, L.C.G., 2006. A dynamic approach to the modelling of correlation credit derivatives using Markov chains. Working Paper. Available at: www.defaultrisk.com.

Hamerle, A., Rösch, D., 2005. Misspecified copulas in credit risk models: How good is gaussian?. Journal of Risk, 8: 41–58.

Herbertsson, A., Rootzén, H., 2006. Pricing k-th-to-default-swaps under default contagion: The matrix-analytic approach. Working Paper. Available at: www.defaultrisk.com.

Hurd, T.R., Kuznetsov, A., 2007. Affine Markov chain model of multifirm credit migration. Journal of Credit Risk, 3: 3–29.

International Financing Review, 2006. Mortality bond issue upsized. International Financing Review, 1656(11/04/2006): 40.

Jacod, J., Shiryaev, A.N., 1987. Limit Theorems for Stochastic Processes. Springer, Berlin, Heidelberg, New York.

Jarrow, R.A., Lando, D., Yu, F., 2005. Default risk and diversification: Theory and empirical implications. Mathematical Finance, 15: 1–26.

Jarrow, R.A., Yu, F., 2001. Counterparty risk and the pricing of defaultable securities. The Journal of Finance, 56: 1765–1799.

Jeffrey, P., 2006. A practitioner's guide to securitisation. City & Financial Publishing, Surrey.

Joe, H., 1997. Multivariate models and dependence concepts. Volume 73 of Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Jorion, P., Shi, C., Zhang, S., 2008. Tightening credit standards: The role of accounting quality. Forthcoming in Review of Accounting Studies.

Joshi, M., Stacey, A., 2006. Intensity gamma: A new approach to pricing portfolio credit derivatives. Risk, July 2006: 78–83.

Kalemanova, A., Schmid, B., Werner., R., 2007. The normal inverse gaussian distribution for synthetic CDO pricing. The Journal of Derivatives, 14: 80–93.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association, 90: 773–795.

Keyfitz, N., 1985. Applied mathematical demography. 2nd Ed., Springer, New York.

Kiesel, R., Scherer, M., 2007. Dynamic credit portfolio modelling in structural models with jumps. Working Paper. Available at: www.defaultrisk.com.

Kristen, R., 2007. Design und Bewertung von Sterblichkeitsderivaten. Diploma Thesis. Ulm University.

Landén, C., 2000. Bond pricing in a hidden Markov model of the short rate. Finance and Stochastics, 4: 371–389.

Lando, D., 1998. On Cox processes and credit risky securities. Review of Derivatives Research, 2: 99–120.

Lando, D., 2004. Credit risk modeling: Theory and applications. Princeton University Press, Princeton.

Lando, D., Nielsen, M.S., 2008. Correlation in corporate defaults: Contagion or conditional independence?. Working Paper. Copenhagen Business School.

Lee, R., 2001. Predicting human longevity. Science, 292: 1654–1655.

Lee, J.-P., Yu, M.-T., 2002. Pricing default-risky CAT bonds with moral hazard and basis risk. The Journal of Risk and Insurance, 69: 25–44.

Löffler, G., 2003. The effects of estimation error on measures of portfolio credit risk. Journal of Banking and Finance, 27: 1427–1453.

Löffler, G., 2007. The complementary nature of ratings and market-based measures of default risk. The Journal of Fixed Income, Summer 2007: 38–47.

Li, D., X., 2000. On default correlation: A copula function approach. The Journal of Fixed Income, March 2000: 43–54.

Lin, Y., Cox, S.H., 2005. Securitization of mortality risks in life annuities. The Journal of Risk and Insurance, 72: 227–252.

Lin, Y., Cox, S.H., 2006. Securitization of catastrophe mortality risks. Working Paper. Youngstown State University and Georgia State University.

Lindskog, F., McNeil, A.J., 2003. Common Poisson shock models: Applications to insurance and credit risk modelling. ASTIN Bulletin 33: 209–238.

Linfoot, A., 2007. Financing catastrophic risk: Mortality bond case study. Presentation on behalf of Scottish Re, 02/14/2007.

Logisch, H.H., 2007. A case study about catastrophe mortality securitization. Diploma Thesis. Ulm University and Commerzbank AG.

Longstaff, F.A., Rajan, A., 2008. An empirical analysis of the pricing of Collateralized Debt Obligations. The Journal of Finance, 63: 529–563.

Luciano, E., Vigna, E., 2005. Non mean reverting affine processes for stochastic mortality. ICER Applied Mathematics Working Paper Series, Working Paper no. 4/2005.

Marshall, A.W., Olkin, I., 1988. Families of multivariate distributions. Journal of the American Statistical Association, 83: 834–841.

Merton, R., 1974. On the pricing of corporate debt: The risk structure of interest rates. The Journal of Finance, 29: 449–470.

Meyer, P.-A., 1971. Démonstration simplifée d'un théorème de Knight. In Séminaire de Probabilités V, Lecture Notes in Mathematics 191, 191–195. Springer, Berlin, Heidelberg, New York.

McNeil, A.J., Frey, R., Embrechts, P., 2005. Quantitative risk management. Princeton Series in Finance. Princeton University Press, Princeton.

Milevsky, M.A., Promislow, S.D., 2001. Mortality derivatives and the option to annuitize. Insurance: Mathematics and Economics, 29: 299–318.

Miltersen, K.R., Persson, S.A., 2005. Is mortality dead? Stochastic force of mortality determined by no arbitrage. Working Paper. University of Bergen.

Moody's Investors Service, 2003. Default & recovery rates of corporate bond issuers. A Statistical Review of Moody's Ratings Performance, 1920-2002.

Moody's Investors Service, 2007. Structured Finance Rating Transitions: 1983-2006.

Mortensen, A., 2006. Semi-analytical valuation of basket credit derivatives in intensity-based models. The Journal of Derivatives, 13: 8–26.

Muermann, A., 2003. Actuarially consistent valuation of catastrophe derivatives. The Wharton Financial Institution Center Working Paper Series, 03-18. University of Pennsylvania.

Nelsen, R.B., 2006. An introduction to copulas. Springer Series in Statistics. Springer, New York.

Niehaus, G., 2002. The allocation of catastrophe risk. Journal of Banking and Finance, 26: 585–596.

Norros, I., 1986. A compensator representation of multivariate life length distributions, with applications. Scandinavian Journal of Statistics, 13: 99–112.

Oeppen, J., Vaupel, J.W., 2002. Broken limits to life expectancy. Science, 269: 1029–1031.

Olivieri, A., 1987. Uncertainty in mortality projections: an actuarial perspective. Insurance: Mathematics and Economics, 29: 231–245.

Olshansky, S.J., Carnes, B.S., Désquelles, A., 2001. Demography: Prospects for human longevity. Science, 291: 1491–1492.

Sircar, R., Papageorgiou, E., 2007. Multiscale intensity models and name grouping for valuation of multi-name credit derivatives. Working Paper. Available at SSRN: http://ssrn.com/abstract=995277.

Petrella, G., 2004. An extension of the Euler Laplace transform inversion algorithm with applications in option pricing. Operations Research Letters, 32: 380–389.

Pitacco, E. , 2004. Longevity risk in living benefits. In: "Developing an annuity market in europe". Fornero, E., and Luciano, E., (Eds.). Edward Elgar, Cheltenham.

Prahl, J., 1999. A fast unbinned test on event clustering in Poisson processes. Working Paper. University of Hamburg.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., 2007. Numerical recipes: The art of scientific computing. Cambridge University Press.

Protter, P.E., 2005. Stochastic integration and differential equations. Springer, Berlin, Heidelberg, New York.

Robert, C. P., Casella, G., 1999. Monte Carlo statistical methods. Springer Texts in Statistics. Springer, New York.

Rogge, E., Schönbucher, P.J., 2003. Modelling dynamic portfolio credit risk. Working Paper. Available at: www.defaultrisk.com.

Schönbucher, P.J., Schubert, D., 2001. Copula dependent default risk in intensity models. Working Paper. Available at: www.defaultrisk.com.

Schönbucher, P.J., 2005. Portfolio losses and the term structure of loss transition rates: A new methodology for the pricing of portfolio credit derivatives. Working Paper. Available at: www.defaultrisk.com.

Schrager, D., 2006. Affine stochastic mortality. Insurance: Mathematics and Economics, 38: 81–97.

Shaked, M., Shanthikumar, J.G., 1987. The multivariate hazard construction. Stochastic Processes and their Applications, 24: 241–258.

Sidenius, J., Piterbarg, V., Andersen, L., 2005. A new framework for dynamic credit portfolio loss modeling. Working Paper. Available at: www.defaultrisk.com.

Singleton, K., 2001. Estimation of affine asset pricing models using the empirical characteristic function. Journal of Econometrics, 102: 111–141.

Smith, A., Spiegelhalter, D., 1982. Bayes factors and choice criteria for linear models. Journal of the Royal Statistical Association, Series B 42: 213–220.

Standard and Poor's, 2006. Ratings raised on Vita Capital II's catastrophe-indexed notes.

Tasche, D., 2007. Validation of internal rating systems and PD estimates. In: "The Analytics of Risk Model Validation". Christodoulakis, G., Satchell, S., (Eds.), Elsevier.

Therneau, T.M., Grambsch, P.M., 2001. Modeling survival data. Extending the Cox model. Statistics for Biology and Heath. Springer, Berlin, Heidelberg, New York.

Vasicek, O., 1991. Limiting loan loss probability distribution. Working Paper. Available at: www.defaultrisk.com.

Walter, U. , 1996. Die Bewertung von Zinsoptionen. Gabler Verlag, Wiesbaden.

Wang, S.S. , 2000. A Class of distortion operations for pricing financial and insurance risks. The Journal of Risk and Insurance, 67: 15–36.

Wang, S.S. , 2002. A universal framework for pricing financial and insurance risks. ASTIN Bulletin, 32: 213–234.

Yoshizawa, Y., 2003. Moody's approach to rating synthetic CDOs. Moody's Investors Service.

Young, V.R., 2004. Pricing in an incomplete market with an affine term structure. Mathematical Finance, 14: 359–381.

Yu, F., 2007. Correlated defaults in intensity-based models. Mathematical Finance, 17: 155–173.

Zusammenfassung

In der vorliegenden Arbeit beschäftigen wir uns mit der Frage, welche Modelle für die Modellierung von strukturierten Finanzprodukten notwendig und geeignet sind. Diese Frage wird aus zwei Richtungen beleuchtet. Einerseits erstellen wir einen Modellrahmen, der viele Modelle aus der Literatur einschließt, und leiten strukturelle Aussagen innerhalb dieses Rahmens her. Andererseits untersuchen wir konkrete Modellspezifikationen im Zusammenhang mit tatsächlichen Daten und können so die Modelle hinsichtlich ihrer Fähigkeit vergleichen, diese Daten zu erklären. Neben einer kurzen Einleitung untergliedert sich die Arbeit in drei Hauptkapitel sowie zwei Anhänge, deren wichtigste Beiträge wir im Folgenden kurz zusammenfassen und diskutieren.

Kapitel 2 bildet die Grundlage für die Anwendungen in den Kapiteln 3 und 4 und beinhaltet die meisten theoretischen Ergebnisse der Arbeit. Wir führen in diesem Kapitel einen Modellrahmen ein, um einen Vektor von Stoppzeiten abzubilden. In den späteren Anwendungen stellen diese Stoppzeiten die Ausfallzeitpunkte von Unternehmen oder die Sterbezeitpunkte von Versicherten dar. Das Besondere an unserem Vorgehen ist, dass unser Ansatz viele Modelle aus der Literatur als Spezialfälle beinhaltet. Dadurch trägt unsere Untersuchung einerseits zu einem tieferen Verständnis dieser Modelle bei, und andererseits helfen unsere Ergebnisse die Vielzahl der Kreditportfolio- und CDO-Modelle in der Literatur zu strukturieren. Beispielsweise zeigen unsere Analysen, dass einige Spezifikationen letztendlich ähnliche oder beinahe identische Modelle mit sich bringen.

Wir beginnen dieses Kapitel mit der Einführung unseres *Stoppzeitmodells* (*stopping times model*), welches grundlegend für alle Fragestellungen dieser Dissertation ist. Das Modell stellt eine flexible „Modell-Plattform" dar mit zahlreichen nützlichen Eigenschaften. So hängen bedingte Überlebenswahrscheinlichkeiten im Modell sowohl von einem *Hintergrundprozess* (*background process*) ab, der sich unabhängig von den Stoppzeiten entwickelt, als auch von den Realisationen der Stoppzeiten selbst. Des Weiteren können Realisationen der Stoppzeiten zusammenfallen. Übersetzt in einen Kreditportfoliokontext bedeutet dies, dass in unserem Modell gemeinsame Ausfälle möglich sind sowie Ansteckungseffekte auftreten können, bei denen Ausfälle im Portfolio die Überlebenswahrscheinlichkeiten der anderen Un-

ternehmen beeinflussen.

Anschließend, im Abschnitt 2.2, beginnen wir mit der Analyse unseres Modellrahmens, beschränken uns dabei zunächst aber nur auf eine einzelne Stoppzeit. In Proposition 2.2.1 leiten wir die Intensität des Prozesses her, der das Auftreten der Stoppzeit beschreibt. Unsere Diskussion zeigt, dass Sprünge des die Stoppzeit ,,auslösenden" Prozesses überflüssig sind, so lange man nur an *einer* Stoppzeit interessiert ist. Basierend auf dem Intensitätsprozess, erhalten wir dann eine Formel für die Einzelüberlebenswahrscheinlichkeiten. Zusätzlich weisen wir auf eine Verbindung zwischen unserem Modell und dem Ansatz aus Collin-Dufresne et al. (2004) für den Fall einer einzelnen Stoppzeit hin.

Im folgenden Abschnitt 2.3 berechnen wir dann gemeinsame Überlebenswahrscheinlichkeiten und die entsprechenden Intensitäten für den ganzen Vektor von Stoppzeiten. Wir zeigen, dass gemeinsame Sprünge der Prozesse, die die Stoppzeiten ,,auslösen", eine ganz entscheidende Rolle für deren Abhängigkeitsstruktur spielen. Es zeigt sich, dass ohne diese Sprünge Realisationen von Stoppzeiten nicht zusammenfallen können. Darüber hinaus stellen wir eine interessante Querverbindung zwischen unserem Modell und den *common Poisson shock models* her, die in Lindskog and McNeil (2003) behandelt werden. Wir illustrieren diese Verbindung mit einem Beispiel und zeigen, dass das *Intensity Gamma* Modell aus Joshi and Stacey (2006) letztendlich ein solches Modell darstellt (Beispiel 2.3.1).

Abschnitt 2.4 legt einen Konstruktions- und Simulationsalgorithmus für unser Stoppzeitmodell vor (Algorithmus 2.4.1). Da die Stoppzeiten in unserem Modellrahmen zusammenfallen können, stellt dieser eine Erweiterung des Algorithmus von Yu (2007) dar.

Aus einem statischen Blickwinkel heraus analysieren wir nachfolgend die modellimplizite Abhängigkeitsstruktur, d.h. wir betrachten Abhängigkeiten zwischen den Überlebensereignissen über einen festen Zeitraum. Wir führen zunächst die wichtigen Begriffe der *bedingten Unabhängigkeit* (*conditional independence*) sowie der Ansteckung (*contagion*) ein und formulieren dann in Proposition 2.5.1 Bedingungen, unter denen unser Stoppzeitmodell der Annahme der bedingten Unabhängigkeit genügt. Ist diese Annahme erfüllt, so kann die modellimplizite Abhängigkeitsstruktur über eine Copula Funktion charakterisiert werden, die bereits in Marshall and Olkin (1988) untersucht wurde. Dies zeigen wir in Satz 2.5.2. Unser Ergebnis bringt mit sich, dass viele Modelle in der Literatur eine schon bekannte Abhängigkeitsstruktur zwischen den Stoppzeiten implizieren.

Des Weiteren diskutieren wir das Clustering der Stoppzeiten in der Zeit, indem wir die Dynamik des *Verlustprozesses* (*loss process*) untersuchen, welcher das Auftreten

der Stoppzeiten zählt. Dadurch liefern wir eine dynamische Charakterisierung der Abhängigkeiten zwischen den Stoppzeiten. In einem ersten Schritt passen wir in Korollar 2.6.1 ein Resultat von Meyer (1971) an unseren Modellrahmen an. Dieses bildet die Grundlage für einen wichtigen statistischen Test, der in Unterabschnitt 3.1.4 betrachtet wird. Anschließend führen wir eine Größe ein, die wir als *erwartete Volatilität* (*expected volatility*) bezeichnen, und schlagen diese als Maß für das Clustering der Stoppzeiten in der Zeit vor. Wir diskutieren dessen Eigenschaften und zeigen seinen Wert für die Analyse von konkreten Modellspezifikationen.

Im Anschluss befassen wir uns mit der Frage der analytischen Handhabbarkeit unseres Modellrahmens. In einem ersten Schritt formulieren wir in Unterabschnitt 2.7.1 Bedingungen, unter denen die charakteristische Funktion eines stochastischen Prozesses semi-analytisch berechnet werden kann, indem man ein System von gewöhnlichen Differentialgleichungen löst. Dies wird dann in Satz 2.7.1 bewiesen. Obwohl wir flexiblere Spezifikationen als in Duffie et al. (2000) und Duffie et al. (2003) zulassen, erhalten wir einen vergleichbaren, hohen Grad an analytischer Handhabbarkeit. In einem zweiten Schritt, in Unterabschnitt 2.7.2, zeigen wir dann, wie unser Stoppzeitmodell angepasst werden muss, damit dieses generelle Ergebnis für die Berechnung von Überlebenswahrscheinlichkeiten und der Verteilung des Verlustprozesses angewandt werden kann. Diese stellen zentrale Größen für die Anwendungen in den Kapiteln 3 und 4 dar.

Wir schließen dieses Kapitel, indem wir eine Brücke zwischen unserem allgemeinen Stoppzeitmodell auf der einen und den spezifischeren Modellen auf der anderen Seite schlagen, die für die Anwendungen betrachtet werden. Darüber hinaus veranschaulichen wir den Nutzen der entwickelten theoretischen Resultate für die Analyse von konkreten Modellspezifikationen in einer beispielhaften Diskussion des Kreditportfoliomodells von Duffie and Gârleanu (2001).

In **Kapitel 3** verknüpfen wir Theorie und Anwendung. Die größte Schwierigkeit bei der Modellierung von strukturierten Kreditprodukten besteht darin, das zugrunde liegende Portfolio und insbesondere seine Abhängigkeitsstruktur richtig zu beschreiben. Fehlspezifikationen in Verbindung mit dieser Abhängigkeitsstruktur gehen auf eine nicht-lineare Weise in die geschätzte Verlustverteilung des strukturierten Kreditprodukts ein. Das Hauptaugenmerk des dritten Kapitels liegt deshalb zum einen auf der Frage, welche Modelle die in tatsächlichen Daten beobachtete Dynamik von Kreditportfolios am besten beschreiben, und zum anderen auf den Konsequenzen der Ergebnisse für die Modellierung von strukturierten Kreditprodukten.

Im ersten Abschnitt dieses Kapitels schätzen wir Ausfallintensitäten für eine große Anzahl von US- und Nicht-US-Unternehmen. Im Gegensatz zu Das et al. (2007)

zeigen wir, dass die geschätzten Intensitäten in der Lage sind, das beobachtete Clustering von Ausfallzeitpunkten in der Zeit zu erklären, obwohl wir unsere Intensitäten basierend auf beobachteten Größen wie der *Expected Default Frequency* (EDF) schätzen und keine zusätzlichen Ansteckungseffekte oder unbeobachtbare Prozesse einführen (Tabelle 3.3). Des Weiteren untersuchen wir die Fähigkeit unseres Modells, Firmen aufgrund ihrer Ausfallwahrscheinlichkeit zu sortieren. Wir stellen fest, dass unser Modell Prognosewerte liefert, die besser sind als jene, die von Duffie et al. (2007) für deren Regressionsmodell und einen ähnlichen Datensatz angegeben werden (Tabelle 3.2).

In Abschnitt 3.2 führen wir dann unser zeitstetiges Modell ein, mit der Absicht, die gemeinsame Dynamik der Ausfallintensitäten abzubilden. Das Modell schließt andere bekannte Modelle in der Literatur mit ein. Im einfachsten Fall folgen die Intensitäten einem *Cox-Ingersoll-Ross*-Prozess (siehe Cox et al. (1985)). Basierend auf unseren theoretischen Ergebnissen aus Kapitel 2 leiten wir eine Formel her (Proposition 3.2.1), die für die Modellumsetzung wichtig ist, und analysieren die Dynamik der Intensitäten in den verschiedenen Modellversionen. Insgesamt bildet das Modell eine hervorragende Grundlage für unsere empirische Analyse, da die ineinander geschachtelten Modellversionen gut verglichen werden können. Insbesondere können wir der Fragestellung nachgehen, ob und wann einfache Modelle ausreichen, um die Ausfallintensitäten treffend zu beschreiben.

Nachdem wir unser Modell eingeführt haben, betrachten wir in Abschnitt 3.3 dessen Kalibrierung auf Einzelfirmenebene. Nach der Entwicklung eines Schätzalgorithmus in Unterabschnitt 3.3.1 vergleichen wir die einzelnen Modellversionen hinsichtlich ihre Fähigkeit, die Ausfallintensitäten jeder einzelnen Firma zu erklären und untersuchen die Ausfallprognosekraft der Modelle. Speziell für Unternehmen niedriger Bonität stellen wir fest, dass Modelle mit Intensitätssprüngen besser in der Lage sind, die Intensitäten abzubilden als Modelle, welche ausschließlich auf Diffusionen basieren (Tabelle 3.7). Allerdings führen komplexere Modelle nicht zu besseren Ergebnissen hinsichtlich der Prognose von Ausfällen (Tabelle 3.8).

Da alle Versionen des betrachteten Intensitätsmodells auf der Annahme von bedingt unabhängigen Ausfällen beruhen, untersuchen wir in Abschnitt 3.4 detailliert die Folgen dieser Annahme. Wir simulieren Ausfalldaten basierend auf einem Modell, in dem Ausfälle die Intensitäten der überlebenden Firmen beeinflussen. Anschließend schätzen wir falsche Modelle – alle basierend auf der Annahme der bedingten Unabhängigkeit – als auch das richtige Modell, das die Daten ursprünglich erzeugt hat, und untersuchen die Fähigkeit der Modelle, die Portfolioverlustverteilung zu prognostizieren. Wir stellen fest, das Schätzfehler einen deutlich größeren Einfluss auf die Ergebnisse besitzen als die Annahme der bedingten Unabhängigkeit (Tabelle 3.9). Obgleich Ansteckungseffekte eine entscheidende Rolle im Daten

erzeugenden Modell gespielt haben, führen Modelle, die auf der bedingten Unabhängigkeit beruhen, zu ähnlichen Ergebnissen wie das geschätzte, richtige Modell (Tabelle 3.10).

Gegen Ende des Kapitels präsentieren wir unsere Modellschätzung auf Portfolioebene. Nach der Einführung eines Schätzalgorithmus in Unterabschnitt 3.5.1, schätzen wir die verschiedenen Modellversionen und vergleichen deren Befähigung, die Portfoliointensität zu erklären. Unter Verwendung der geschätzten Parameter simulieren wir dann Pfade der Portfoliointensität und des zugehörigen Portfolioverlustes. Auf diese Weise können wir Zeitreihen von Ratings für hypothetische, strukturierte Kreditprodukte berechnen, denen das simulierte Portfolio zugrunde liegt, und erhalten schlussendlich dadurch Rating-Transitionsmatrizen. Wir stellen fest, dass basierend auf dem betrachteten Datensatz einfache und komplexe Modelle zu ähnlichen Risikoprofilen für strukturierte Kreditprodukte führen (Tabelle 3.17) und dass generelle Größen, wie die durchschnittliche Ausfallrate im zugrunde liegenden Portfolio, einen wesentlich größeren Einfluss auf die Ergebnisse besitzen als die Modellwahl.

Kapitel 4 stellt zuletzt die zweite Anwendung unseres Stoppzeitmodells für die Analyse von mortalitätsbedingten Katastrophenbonds vor.

Da wir in unserer Untersuchung konkrete Transaktionen betrachten, beginnen wir in Abschnitt 4.1 mit einem kurz gefassten Überblick über den Markt für diese Wertpapiere und beschreiben die Transaktionen, die untersucht werden.

Darauf folgt in Abschnitt 4.2 die Einführung unseres Modells für die Analyse und Bewertung von mortalitätsbedingten Katastrophenbonds. Die Modellspezifikation beinhaltet zwei Komponenten: Eine *Grundkomponente* (*Baseline Component*), die einer Diffusionsspezifikation folgt und die normale Entwicklung der Sterblichkeit über die Zeit beschreibt, und eine *Katastrophenkomponente*, die von einem Nicht-Gausschen Ornstein-Uhlenbeck Prozess getrieben wird. Unseres Wissens nach stellt dieses Modell den ersten vollständig dynamischen Ansatz in der Literatur dar, diese Wertpapiere zu modellieren. Unsere Diskussion des Modells zeigt, dass es vielversprechende Eigenschaften besitzt. Insbesondere können Überlebenswahrscheinlichkeiten semi-analytisch berechnet werden und darauf basierend Versicherungsprämien oder -leistungen bestimmt werden.

Die Schätzung unseres Modells erfolgt in Abschnitt 4.3. Wir diskutieren dort drei verschiedene Kalibrierungsarten für unser Modell: In Unterabschnitt 4.3.1 betrachten wir zunächst dessen Kalibrierung basierend auf historischen Sterblichkeitsdaten. Wir stellen fest, dass insbesondere die Parameter der Katastrophenkomponente großen Unsicherheiten unterliegen. In einem zweiten Schritt schätzen

wir risiko-adjustierte Parametrisierungen des Modells sowohl basierend auf Versicherungspreisen als auch auf Marktpreisen von mortalitätsbedingten Katastrophenbonds.

Im Anschluss daran berechnen wir in Abschnitt 4.4 Risikoprofile und Spread Level für die betrachteten Wertpapiere unter Verwendung der Parametrisierungen aus Abschnitt 4.3. Wir vergleichen unsere Ergebnisse mit denen von so genannten *Risikomodellierungsfirmen* (*risk modeling firms*), auf denen die Entscheidungen von Ratingagenturen und Investoren maßgeblich beruhen. Insgesamt stellen wir fest, dass die Profile großen Unsicherheiten ausgesetzt sind in Bezug auf die zugrunde liegenden Daten und deshalb von allen Marktteilnehmern mit Vorsicht betrachtet werden sollten. Insbesondere sind die Risikokennzahlen der Risikomodellierungsfirmen niedriger als unsere, obwohl keine strukturellen Unterschiede in den Ergebnissen vorliegen (Tabelle 4.9). Dies deutet darauf hin, dass die zugrunde liegenden Annahmen der Risikomodellierungsfirmen eher ,,optimistisch" sind. Durch die Analyse der erhaltenen risiko-adjustierten Parametrisierungen können wir darüber hinaus eine Erklärung für das schnelle Wachstum des Marktes für mortalitätsbedingte Katastrophenbonds in den letzten Jahren geben.

Anhang A präsentiert eine geschlossene Formel für eine wichtige Transformation einer so genannten *basic affine jump diffusion*, die auf Duffie and Gârleanu (2001) zurückgeht. Außerdem leiten wir einen exakten Simulationsprozess (Algorithmus A.0.1) für dieses Modell her. Dieser erweitert den Algorithmus, der für einen Spezialfall des Modells – einen Cox-Ingersoll-Ross-Prozess – bekannt ist und den man in Glasserman (2004) auf Seite 124 findet. Neben diesem Algorithmus liefern wir weitere Details zu unserer Implementierung der Modelle aus Kapitel 3.

Im **Anhang B** zeigen wir, wie man im Modell aus Kapitel 4 die approximative Verteilung des *kombinierten Sterblichkeitsindex* (*combined mortality index*) berechnet. Weitere Graphiken mit Parametersensitivitäten werden für dieses Modell präsentiert.

Danksagung

Bei der Erstellung einer so umfangreichen Arbeit wie einer Dissertation ist man immer auf den Rat, die Hilfe und Unterstützung anderer Menschen angewiesen, bei denen ich mich an dieser Stelle bedanken möchte.

Ich bedanke mich bei meinen Mitstipendiaten im Graduiertenkolleg für zahlreiche Diskussionen und viele schöne Erinnerungen. Insbesondere gilt mein Dank Daniel Bauer und Christian Schmidt nicht nur für fachliche Diskussionen, sondern vor allem auch für viele unterhaltsame Momente. Bei beiden sowie bei Eduard Hergenreider und Helge Scheffler bedanke ich mich außerdem für das Korrekturlesen dieser Arbeit und für wertvolle Hinweise.

Darüber hinaus gilt mein Dank Herrn Prof. Dr. Spodarev und Herrn Prof. Dr. Urban für ihre Zusage als Wahlmitglieder der Prüfungskommission zu fungieren. Insbesondere danke ich auch Herrn Prof. Dr. Urban und der Abteilung Numerik für die in den letzten Jahren zur Verfügung gestellten numerischen Ressourcen, die bei vielen Berechnungen dieser Dissertation zum Einsatz kamen. Mein besonderer Dank gilt auch Herrn Prof. David Lando von der Copenhagen Business School für die erfahrene Gastfreundschaft und zahlreiche hilfreiche Anmerkungen und Hinweise während meines Forschungsaufenthaltes in Dänemark. Besonders möchte ich mich auch bei Prof. Dr. Rüdiger Kiesel für die Übernahme des Zweitgutachtens bedanken und vor allem auch für sein stets offenes Ohr für meine Fragen und Belange. Mein ganz besonderer Dank gilt meinem Doktorvater Prof. Dr. Löffler für zahlreiche Betreuungsgespräche und für die in jeder Hinsicht sehr große Unterstützung in den letzten Jahren. Seine inspirierende Art und eine offene Arbeitsatmosphäre bildeten die Grundlage für eine interessante und angenehme Zusammenarbeit.

Ganz besonders möchte ich mich natürlich auch bei meiner Familie für Ihre immerwährende Unterstützung bedanken. Für dies und für so vieles mehr danke ich meinen Eltern Gert und Waltraud, meiner Lebensgefährtin Carolin sowie meinen Geschwistern Peter, Maria, Thomas, Karin, Ingrid, Stephan und Anna.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Es wurden keine außer den angegebenen Quellen verwendet und die aus diesen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.
Ferner erkläre ich, dass die vorliegende Arbeit in keinem anderen Promotionsverfahren und an keiner anderen Stelle als Prüfungsleistung verwendet wurde.

Ulm, den 30. September 2008 _____

(Florian Kramer)