Some asymptotic results on non-standard likelihood ratio tests, and Cox process modeling in finance

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Alexander Szimayer

aus

Mosbach (Baden)

Bonn 2002

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

1. Referent:	Prof. Dr. Manfred Schäl
2. Referent:	Prof. Dr. Sergio Albeverio
Tag der Promotion:	17. September 2002

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Dr. Manfred Schäl (Bonn University) who guided me through various fields of stochastic analysis, especially mathematical finance, and who often gave me valuable advice and support. Also I am grateful to Dr. Angelika May for creating a productive research environment in her Financial Engineering Group in Research Center caesar, Bonn. I am indebted to Prof. Dr. Claudia Klüppelberg (Munich University of Technology) for encouraging me to study non-standard aspects of the theory of statistical testing. I also take pleasure in thanking Prof. Dr. Ross Maller (University of Western Australia) for his sound advice and patient explanations. Finally, I would like to thank my co-workers Dr. Milan Borkovec, Dr. Mark Van De Vyver, Dr. Niklas Wagner, Taras Beletski, Gabriel Frahm, Anke Gleisberg, Marcus Haas, Jörg Hagspiel, and Markus Junker.

I would like to thank the Bundesministerium für Bildung und Forschung for financial support through the program BMBF 03MAM6CA.

Contents

Chapter 1

Introduction

This dissertation consists of two parts. In the first part, the subject of hypothesis testing is addressed. Here, non-standard formulations of the null hypothesis are discussed, e.g., non-stationarity under the null, and boundary hypotheses. In the second part, stochastic models for financial markets are developed and studied. Particular emphasis is placed on the application of Cox processes.

Part one begins with a survey of time-series models which allow for conditional heteroscedasticity and autoregression, AR-GARCH models. These models reduce to a white noise model, when some of the conditional heteroscedasticity parameters take their boundary value at zero, and the autoregressive component is in fact not present. The asymptotic distribution of the pseudo-log-likelihood ratio statistics for testing the presence of conditional heteroscedasticity and the autoregression term is reproduced, see Andrews (1999b), and Klüppelberg et al. (2002). These results are applied to financial market data. The model parameters are estimated and tests for the reduction to white noise are performed. The impact of these results on risk measurement is discussed by comparing several Value-at-Risk calculations assuming the alternative model specifications. Furthermore, the power function of these tests is examined by a simulation study of the ARCH(1) and the AR(1)-ARCH(1) models. First, the simulations are carried out assuming Gaussian innovations and then, the Gaussian distribution is replaced by the heavy tailed t-distribution. This reveals that a substantial loss of power is associated with the use of heavy tailed innovations. The implications of these results on financial time-series modeling is shown in the context of Value-at-Risk (VaR) calculation. Using a sample size of 500 observations, we show that in most cases no significant conditional heteroscedasticity effects are found, i.e. the empirical LR statistics suggests rejecting the null hypothesis of white noise, but not with sufficient power.

A related testing problem arises in the analysis of the Ornstein-Uhlenbeck (OU) model, driven by Levy processes. This model is designed to capture mean reverting behaviour if it exists; but the data may in fact be adequately described by a pure Levy process with no OU (autoregressive) effect. For an appropriate discretized version of the model, likelihood methods are utilized to test for such a reduction of the OU process to Levy motion, deriving the distribution of the relevant pseudo-log-likelihood ratio statistics, asymptotically, both for a refining sequence of partitions on a fixed time interval with mesh size tending to zero, and as the length of the observation window grows large. These analyses are non-standard in that the mean reversion parameter vanishes under the null of a pure Levy process for the data. Despite this a very general analysis is conducted with no technical restrictions on the underlying processes or parameter sets, other than a finite variance assumption for the Levy process. As a special case, for Brownian Motion as driving process, the limiting distribution is deduced in a quite explicit way, finding results which generalise the well-known Dickey-Fuller ("unit-root") theory.

Part two of this dissertation considers the application of Cox processes in mathematical finance. Here, we discuss a framework for the valuation of employee share options (ESO), and credit risk modeling. One popular approach for ESO valuation involves a modification of standard option pricing models, augmenting them by the possibility of departure of the executive at an exogenously given random time, see Carr and Linetsky (2000). Such models are called reduced form models, in contrast to structural models that require measures of the employee's utility function and other unobservable quantities. Here, an extension of the reduced form model for the valuation of ESOs is developed. This model incorporates and emphasises employee departure, company takeover, performance vesting and other exotic provisions specific to ESOs. The fundamental components of the setup are the financial market carrying the relevant tradable assets and two random times announcing employee departure and takeover, where the two random times can both be associated with the first jumps of two different Cox processes. By the nature of the construction, the market model is incomplete. This market incompleteness results in a set of pricing systems, i.e. equivalent martingale measures, rather than a single price for a given contingent claim. For stereotypical ESOs the range of possible fair values is given. In addition, the prices of these ESOs are evaluated under several prominent martingale measures. Furthermore, possible limitations of the proposed model are explored by examining departures from the crucial assumptions of no-arbitrage, e.g., by considering the effects of insider information.

In a continuous time market model, credit risk modeling and pricing of credit derivatives is discussed. In the approach we adopt, credit risk is described by the interest rate spread between a corporate bond and a government bond. This spread is modeled in terms of explaining variables. For this purpose, a specific market model consisting of four assets is considered where the default process of the company is incorporated in a risky money market by a Cox process, see Lando (1998). We show that this market model has a unique equivalent martingale measure and is complete. As a consequence, contingent claim valuation can be executed in the usual way. This is illustrated with the valuation of a convertible bond which fits naturally in the given setting.

Chapter 2

Testing for Conditional Heteroscedasticity

2.1 Introduction

Conditional heteroscedasticity models introduced by Engle (1982) are well established and frequently applied to time-series. Generalizations of these so called ARCH-models exist in various ways, see for example Bollerslev (1986) and Bera and Higgins (1993) for the GARCH specification. In the field of economics conditional heteroscedasticity models are of importance, especially for financial time-series. Empirical evidence for conditional heteroscedasticity effects is given by Bollerslev et al. (1992) and moreover, Duan (1995) develops an GARCH option pricing model. The progression of the GARCH models is also reflected in risk management where traditionally Gaussian white noise models are applied to describe financial time series. The choice of a white noise model seems quite appealing since this setup is a discrete time version of the classical Black&Scholes model that allows also for measuring the risk of derivative securities, see Hull (1993). In recent years, conditional heteroscedasticity models received growing attention for risk management, see Jorion (2001), and also Frey and McNeil (2000) who apply the GARCH framework to Value-at-Risk calculation.

Overall, conditional heteroscedasticity models incorporate the white noise model as a special case, i.e. when the conditional heteroscedasticity degenerates to the homoscedastic case. And therefore, these models provide a more general framework than the white noise specification. But the generality of conditional heteroscedasticity models demands more sophisticated methods, e.g., in Value-at-Risk calculation and option pricing and of course, more computational effort. For this reason, testing conditional heteroscedasticity models for reduction to the white noise model/conditional homoscedasticity becomes an important subject. Unfortunately, the problem of testing for homoscedasticity in GARCH type models can not be covered by standard theory, tests like Lagrange multiplier (LM) test and likelihood ratio (LR) test in their general form fail. The reason for this is that the null hypothesis of conditional homoscedasticity corresponds to a boundary value of the parameter space with respect to the general model. Nevertheless, this topic has been studied in recent years and results on various conditional heteroscedasticity models have been established, see Demos and Sentana (1998), Andrews (1999b), and Klüppelberg et al. (2002).

Here, we survey time-series models allowing for conditional heteroscedasticity and autoregression. In particular, we study the ARCH(1), GARCH(1,1), and AR(1)-GARCH(1,1) model. These models reduce to white noise, i.e. the Black&Scholes model, when some of the conditional heteroscedasticity parameters take their boundary value at zero, and the autoregressive component is in fact not present. We state the asymptotic distribution of pseudo-log likelihood ratio statistics for testing the presented conditional heteroscedasticity models for reduction to white noise. The theoretical results studied here are applied to financial data, i.e. log-returns of stock prices. We estimate the model parameters and further on, we test on reduction to white noise. The empirical observations indicate whether the time-series exhibits conditional heteroscedasticity or the data corresponds to white noise. We show examples where the test accepts the model reduction and hence, the more feasible Black&Scholes framework is sufficient. The impact of these results on risk measurement is discussed by comparing Value-at-Risk calculations under alternative model specifications, i.e. the conditional heteroscedasticity model and the Black&Scholes approach.

Furthermore, we study the power function of the LR test on conditional heteroscedasticity what is done for the ARCH(1), and AR(1)-ARCH(1) model specification. Under the null hypothesis the asymptotic distribution of LR statistics is given in a closed form expression that is tractable for calculations, whereas under the alternative we have to conduct a simulation study to attain the distribution function. The simulations on the alternative are carried out where we primarily use Gaussian innovations. For the ARCH(1) model, we also investigate the impact of heavy tailed innovations on the power function, and we find a loss of power compared to the Gaussian case. Extending the model for an autoregressive component of order one, we obtain the AR(1)-ARCH(1) model, that is widely used for describing financial time-series, especially in the context of Value-at-Risk calculation. For log-return series from the German and US equity market and the standard VaR-sample size of 500 days, we show that in most cases we are not able to find significant conditional heteroscedasticity effects, i.e. the empirical LR statistics suggest to reject the null hypothesis of white noise, but not at suitable power. This conclusion becomes even more distinct in the presence of heavy tailed innovations what is one of the so-called "stylized facts" we know about financial data.

2.2 AR-GARCH Models and the LR Statistics

In this section, time-series models allowing for conditional heteroscedasticity and autoregression are presented. Additionally, for each model, we reproduce the form of the asymptotic distribution of the likelihood ratio (LR) statistics for testing on reduction to white noise. This is carried out for an AR-GARCH model studied by Klüppelberg et al. (2002), and the well known ARCH and GARCH models (see Bera and Higgins, 1993, and Bollerslev, 1986), where in the two latter cases testing on conditional homoscedasticity is discussed in Demos and Sentana (1998) and Andrews (1999b).

First, we specify the probabilistic setting where the time-series is placed in, and recall the form of the maximum likelihood estimator and the deviance for testing on model reduction in a general framework. Let $(\Omega, P, \mathcal{F}, (\mathcal{F}_t)_{t\geq 1})$ be a filtered probability space in discrete time. The innovations driving the time-series are given by an i.i.d. family of random variables $(\varepsilon_t)_{t\geq 2}$ with zero expectation and unit variance, and a finite fourth moment. The filtration is given by

$$\mathcal{F}_t = \mathcal{F}\left(\varepsilon_2, ..., \varepsilon_t\right), \quad \text{for } t \geq 2, \tag{2.1}$$

and $\mathcal{F}_1 = \{\emptyset, \Omega\}$.

The time-series $(X_t)_{t\geq 1}$ with initial value $X_1 \in \mathbb{R}$ is defined by

$$X_t = \alpha X_{t-1} + \sigma_t \varepsilon_t, \qquad \text{for } t = 2, 3..., \tag{2.2}$$

where $\alpha \in [-1, 1]$ and $(\sigma_t)_{t\geq 2}$ is a positive predictable process. With $e_t = \sigma_t \varepsilon_t$ for $t \geq 2$, we can write Equation (2.2) in the form

$$X_t = \alpha X_{t-1} + e_t, \qquad \text{for } t = 2, 3... . \tag{2.3}$$

Thus, the process $(X_t)_{t\geq 1}$ is autoregressive with innovations $(e_t)_{t\geq 2}$ showing conditional variance $\mathbb{E}\left\{e_t^2 \,|\mathcal{F}_{t-1}\right\} = \sigma_t^2$, for $t \geq 2$.

The pseudo-log likelihood function for a finite sample of length $T \in \mathbb{N}$ is given by

$$\mathcal{L}_T(\theta) = -\frac{1}{2}\sum_{t=2}^{T} \ln \sigma_t^2 - \frac{1}{2}\sum_{t=2}^{T} \varepsilon_t^2 - \frac{1}{2}(T-1)\ln(2\pi), \tag{2.4}$$

where θ is a vector describing the model, hence θ parameterizes σ_t and ε_t in Equation (2.4).

For testing purposes, we assume that the *true model* is given by $\theta \in \Omega$. For a given subset Ω_H of Ω, we can test the null hypothesis $\theta \in \Omega_H$ versus the alternative $\theta \in \Omega \setminus \Omega_H$. The test utilized in here is the likelihood ratio test, therefore, we define the deviance

$$d_T = -2\left(\mathcal{L}_T(\hat{\theta}_0) - \mathcal{L}_T(\hat{\theta})\right), \tag{2.5}$$

where $\hat{\theta}_0 \in \Omega_H$ is the maximum likelihood estimator for the null hypothesis, and $\hat{\theta}$ is the corresponding estimator for the alternative $\Omega \setminus \Omega_H$. Later on, we specify the conditional heteroscedasticity models and reproduce the asymptotic distribution of the deviance statistics. It is worth mentioning that the asymptotic distribution of the deviance statistics is a non-trivial mathematical task. Testing on conditional homoscedasticity transfers to the problem of testing a boundary hypothesis, since the conditional heteroscedasticity parameters take their boundary value at zero in the conditional homoscedastic case.

Klüppelberg et al. (2002) discuss conditional heteroscedasticity models allowing also for autoregression as given in Equation (2.2) by specifying a AR(1)-GARCH(1,1) model. The conditional variance of the innovations is determined by

$$\sigma_t^2 = \omega + \lambda e_{t-1}^2 + \delta\sigma_{t-1}^2 = \omega + \lambda\sigma_{t-1}^2\varepsilon_{t-1}^2 + \delta\sigma_{t-1}^2, \qquad \text{for } t = 2, 3..., \tag{2.6}$$

where $\omega > 0$ and $\lambda \geq 0$ and $\delta \geq 0$, and $\theta = (\alpha, \omega, \lambda, \delta) \in \Omega = [-1, 1] \times \mathbb{R}^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+$. The log likelihood function reads as in Equation (2.4) with

$$\sigma_t^2 = \vartheta + \delta^{t-1}(\sigma_1^2 - \vartheta) + \lambda \sum_{i=1}^{t-1} \delta^{i-1} \left(X_{t-i} - \alpha X_{t-1-i}\right)^2 , \quad \text{and} \tag{2.7}$$

$$\varepsilon_t^2 = (X_t - \alpha X_{t-1})^2 \left(\vartheta + \delta^{t-1}(\sigma_1^2 - \vartheta) + \lambda \sum_{i=1}^{t-1} \delta^{i-1} \left(X_{t-i} - \alpha X_{t-1-i}\right)^2 \right)^{-1} \tag{2.8}$$

for $t = 3, ...$, where $\vartheta = \frac{\omega}{1-\delta}$, and $\sigma_1 > 0$, and $\sigma_2^2 = \omega + \delta\sigma_1^2$ and $\varepsilon_2^2 = \frac{(X_2 - \alpha X_1)^2}{\omega + \delta\sigma_1^2}$.

The null hypothesis of conditional homoscedasticity and the absence of autoregression is given by the set $\Omega_H = \{0\} \times \mathbb{R}^+ \times \{0\} \times \{0\}$, i.e. $\alpha = \lambda = \delta = 0$. For $\theta \in \Omega_H$, the log likelihood function stated in Equations (2.7) and (2.8) simplifies to

$$\sigma_t^2 = \omega \quad \text{and} \quad \varepsilon_t^2 = \frac{X_t^2}{\omega}, \qquad \text{for } t = 2, 3..., \tag{2.9}$$

Klüppelberg et al. (2002) computed the asymptotic distribution of the deviance statistics for testing the null hypothesis Ω_H versus the alternative $\Omega \setminus \Omega_H$.

Theorem 2.1 (Klüppelberg et al., 2002) *In the present setting, let μ_3 and μ_4 denote the third and the fourth moment of the innovations (ε_t), and $\mu_3, \mu_4 < \infty$. Then under the null $\mathbf{H}_0 : \alpha = \lambda = \delta = 0$, i.e. $\theta_0 \in \Omega_H$*

$$d_T \xrightarrow{D} N^2 + Z^2 \mathbf{1}_{\{Z \geq 0\}}, \quad \textit{for } T \to \infty, \tag{2.10}$$

where

$$Z = \frac{\mu_3^2}{\sqrt{2(\mu_4 - 1)}} N + \sqrt{\frac{(\mu_4 - 1)^2 - \mu_3^4}{2(\mu_4 - 1)}}\, \widetilde{N}$$

with N and $\widetilde{N}$ independent standard normal random variables.

Remark. (1) In the situation of the theorem, assuming Gaussian innovations implies $\mu_3 = 0$ and $\mu_4 = 3$, and hence

$$d_T \xrightarrow{D} N^2 + \widetilde{N}^2 \mathbf{1}_{\{\widetilde{N} \geq 0\}}, \quad \text{for } T \to \infty, \tag{2.11}$$

where N and $\widetilde{N}$ are again independent standard normal random variables.

(2) The result of Theorem 2.1 remains valid, if we restrict the alternative/full model to the AR(1)-ARCH(1) specification, i.e. $\Omega = [-1, 1] \times \mathbb{R}^+ \times \mathbb{R}_0^+ \times \{0\}$. Then for the null hypothesis of no conditional heteroscedasticity given by $\Omega_H = \{0\} \times \mathbb{R}^+ \times \{0\} \times \{0\}$, the asymptotic distribution of the deviance is given by the righthand side of Equation (2.10).

We point out that the hypothesis of conditional homoscedasticity is formulated by $\lambda = 0$ and $\delta = 0$. However, if $\lambda = 0$, the variance process $(\sigma_t^2)_{t \geq 1}$ is a deterministic function converging to $\vartheta = \frac{\omega}{1-\delta}$. For sufficiently large sample size, conditional homoscedasticity appears though $\delta > 0$. Accordingly, conditional homoscedasticity can be specified by $\lambda = 0$ and $\delta \geq 0$. This causes a nuisance parameter δ to appear that cannot be identified under the null hypothesis. Andrews (1999b) covers this problem for GARCH(1,1). In case of the AR(1)-GARCH(1,1) model, we restrict ourselves to the approach considered by Klüppelberg et al. (2002), hence the hypothesis of conditional homoscedasticity is given by $\lambda = 0$ and $\delta = 0$.

An approach frequently applied for modeling financial time-series is a GARCH(1,1) model, see Bollerslev (1986). Within this framework, the conditional heteroscedasticity is specified, but no autoregression is taken into account, hence $\alpha = 0$, and Equation (2.2) reduces to $X_t = \sigma_t \varepsilon_t$, for $t \geq 2$. The conditional heteroscedasticity is given by

$$\sigma_t^2 = \omega + \lambda e_{t-1}^2 + \delta \sigma_{t-1}^2, \qquad \text{for } t = 2, 3..., \tag{2.12}$$

where $\sigma_1^2 > 0$ is given, and $\omega > 0$, $\lambda \geq 0$, and $\delta \geq 0$, and $\theta = (\omega, \lambda, \delta) \in \Omega = \mathbb{R}^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+$. By setting $\vartheta = \frac{\omega}{1-\delta}$, the log likelihood function in Equation (2.4) is determined by

$$\sigma_t^2 = \vartheta + \delta^{t-1}(\sigma_1^2 - \vartheta) + \lambda \sum_{i=1}^{t-1} \delta^{i-1} X_{t-i}^2, \quad \text{and} \tag{2.13}$$

$$\varepsilon_t^2 = X_t^2 \left(\vartheta + \delta^{t-1}(\sigma_1^2 - \vartheta) + \lambda \sum_{i=1}^{t-1} \delta^{i-1} X_{t-i}^2 \right)^{-1}, \quad \text{for } t = 2, ..., \tag{2.14}$$

Following the approach we presented for the AR(1)-GARCH(1,1) model, the null hypothesis of conditional homoscedasticity could read $\lambda = 0$ and $\delta = 0$. As mentioned before, the parameter δ appears to be a nuisance parameter for this formulation of the null hypothesis, since we can not identify λ and δ simultaneously under the null. Andrews (1999b) shows a way to control this problem. In his framework, he applies stationarity arguments and therefore, he assumes $\delta \in \Delta$ apriori, where $\Delta = [0, \delta_u]$ with $\delta_u < 1$. With this assumption, the parameter space Ω is of the form $\Omega = \mathbb{R}^+ \times \mathbb{R}_0^+ \times \Delta$. Furthermore, he formulates the null hypothesis of conditional homoscedasticity by $\lambda = 0$, hence $\Omega_H = \mathbb{R}^+ \times \{0\} \times \Delta$.

On the parameter space Ω describing the alternative, the information matrix becomes singular under the null hypothesis, hence we cannot identify λ and δ simultaneously. Andrews (1999b) overcomes this problem by fixing $\delta \in \Delta$ in a first step, i.e. the parameter space is restricted to $\Omega_\delta = \mathbb{R}^+ \times \mathbb{R}_0^+ \times \{\delta\}$, for each $\delta \in \Delta$. On each restricted space Ω_δ, a maximum likelihood estimation is carried out, what is possible, since δ is fixed. This results into $\mathcal{L}_T(\hat{\theta}_\delta; \delta)$, where $\hat{\theta}_\delta$ is the maximizer of the log likelihood function on Ω_δ. In a second step, the supremum is taken over all $\delta \in \Delta$, and Equation (3.9) becomes

$$\mathcal{L}_T(\hat{\theta}) = \sup_{\delta \in \Delta} \mathcal{L}_T(\hat{\theta}_\delta; \delta), \tag{2.15}$$

where $\hat{\theta}$ is the maximizing argument that needs not to be unique. When the initial condition is $\sigma_1 = \vartheta = \frac{\omega}{1-\delta}$, the log likelihood does not depend on δ for any $\theta \in \Omega_H$. Hence, the estimator on the hypothesis Ω_H is still given by Equation (2.9) and does not depend on the nuisance parameter δ, at least asymptotically, for large T and arbitrary $\delta \in \Delta$. With this specification, Andrews (1999b) obtains the asymptotic distribution of the deviance statistics d_T under the null hypothesis.

$$d_T \xrightarrow{D} \frac{\mu_4 - 1}{2} \sup_{\delta\in\Delta} Z_\delta^2 \mathbf{1}_{\{Z_\delta \geq 0\}}, \quad \text{for } T \to \infty, \tag{2.16}$$

where μ_4 is the fourth moment of the innovations and $(Z_\delta)_{\delta\in\Delta}$ is a Gaussian process with covariance structure

$$\mathrm{cov}(Z_{\delta_1}, Z_{\delta_2}) = \frac{(1-\delta_1^2)(1-\delta_2^2)}{1-\delta_1\delta_2}, \quad \text{for } \delta_1, \delta_2 \in \Delta.$$

For computational purposes, we can write Equation (2.16) as

$$d_T \xrightarrow{D} c \sup_{\delta\in\Delta} Y_\delta^2 \mathbf{1}_{\{Y_\delta \geq 0\}}, \quad \text{for } T \to \infty, \tag{2.17}$$

where $c = \frac{\mu_4 - 1}{2}$ and $Y_\delta = \sqrt{1-\delta^2} \sum_{i=0}^{\infty} \delta^i \widetilde{Z}_i$, with $\left(\widetilde{Z}_i\right)_{i\geq 0}$ are *iid* standard normal random variables. Furthermore, we can replace c by the estimator $\hat{c}_T$, where

$$\hat{c}_T = \frac{1}{2}\left(\frac{\frac{1}{T}\sum_{t=1}^T X_t^4}{\left(\frac{1}{T}\sum_{t=1}^T X_t^2\right)^2} - 1\right), \tag{2.18}$$

and define a rescaled test statistics

$$\frac{d_T}{\hat{c}_T} \xrightarrow{D} \sup_{\delta\in\Delta} Y_\delta^2 \mathbf{1}_{\{Y_\delta \geq 0\}}, \quad \text{for } T \to \infty, \tag{2.19}$$

where the asymptotic distribution under the null hypothesis is preserved.

Theorem 2.2 (Andrews, 1999b) *In the present setting, let the fourth moment of the innovations (ε_t) be finite. Then under the null $\mathbf{H}_0 : \lambda = 0$, i.e. $\theta_0 \in \Omega_H$*

$$\frac{d_T}{\hat{c}_T} \xrightarrow{D} \sup_{\delta\in\Delta} Y_\delta^2 \mathbf{1}_{\{Y_\delta \geq 0\}}, \quad \textit{for } T \to \infty, \tag{2.20}$$

where $\hat{c}_T$ is given by Equation (2.18), and $Y_\delta = \sqrt{1-\delta^2} \sum_{i=0}^{\infty} \delta^i \widetilde{Z}_i$, with $\left(\widetilde{Z}_i\right)_{i\geq 0}$ iid *standard normal random variables.*

Andrews (1999b) generates the asymptotic critical values by simulation. For $\Delta = [0, 0.95]$, for significance levels 90%, 95%, and 99%, the critical values are 3.06, 4.33, and 7.30 respectively.

Finally, we consider the ARCH(1) model. Properties of this model, and estimation and testing are surveyed in Bera and Higgins (1993). The conditional variance is specified by

$$\sigma_t^2 = \omega + \lambda e_{t-1}^2, \quad \text{for } t = 2, 3..., \tag{2.21}$$

where $\omega > 0$ and $\lambda \geq 0$, and $\theta = (\beta, \lambda) \in \Omega = \mathbb{R}^+ \times \mathbb{R}_0^+$. The log likelihood function in Equation (2.4) is determined by

$$\sigma_t^2 = \beta + \lambda X_{t-1}^2, \quad \text{and} \tag{2.22}$$

$$\varepsilon_t^2 = \frac{X_t^2}{\beta + \lambda X_{t-1}^2}, \quad \text{for } t = 2, \dots . \tag{2.23}$$

The null hypothesis of conditional homoscedasticity is given by $\lambda = 0$, or $\Omega_H = \mathbb{R}^+ \times \{0\}$. The maximum likelihood estimator on the null hypothesis is given by Equation (2.9). The asymptotic distribution of the deviance for testing τ versus Ω is deduced in Demos and Sentana (1998) for Gaussian innovations.

$$d_T \xrightarrow{D} N^2 \mathbf{1}_{\{N \geq 0\}}, \quad \text{for } T \to \infty, \tag{2.24}$$

where N is standard normal. This result can be generalized for non Gaussian innovations by setting $\Delta = \{0\}$ in the GARCH(1,1) model, see Equation (2.16), and hence Equation (2.24) becomes

$$d_T \xrightarrow{D} \frac{\mu_4 - 1}{2} N^2 \mathbf{1}_{\{N \geq 0\}}, \quad \text{for } T \to \infty. \tag{2.25}$$

Applying the result of Andrews (1999b) for the case $\Delta = 0$, we can rescale the deviance by $c_T = \frac{\mu_4 - 1}{2}$. The estimate of c_T is given by $\hat{c}_T$ in Equation (2.18).

Theorem 2.3 (Andrews, 1999/Demos and Sentana, 1998) *In the present setting, let the fourth moment of the innovations* (ε_t) *be finite. Then under the null* $\mathbf{H}_0 : \lambda = 0$, *i.e.* $\theta_0 \in \Omega_H$

$$\frac{d_T}{\hat{c}_T} \xrightarrow{D} N^2 \mathbf{1}_{\{N \geq 0\}}, \quad \textit{for } T \to \infty, \tag{2.26}$$

where $\hat{c}_T$ *is given by Equation (2.18).*

According to Demos and Sentana (1998) the critical values of 1.64, 2.71, and 5.41 are corresponding to the significance levels of 90%, 95%, and 99% respectively.

Alternative	90 %	95%	99 %
ARCH(1) (d^1)	1.64	2.71	5.41
GARCH(1,1) (d^2)	3.06	4.33	7.30
AR(1)-ARCH(1)	3.80	5.13	8.28
AR(1)-GARCH(1,1) (d^3)	3.80	5.13	8.28

Table 2.1: Critical values for testing the Black&Scholes model vs. various alternatives.

Table 2.1 summarizes the critical values for all model specifications presented in this section. Additionally, we indicated by superscript numbers attached to deviances which model is used, and this notation is applied for the subsequent testing procedure. We remark that the statistics d^1 and d^2 are rescaled statistics, hence the asymptotic distribution does not depend on the characteristics of the innovations. Unfortunately, the LR statistics d^3 for testing white noise versus the AR(1)-GARCH(1,1) alternative cannot be rescaled. The critical values depend on the third and fourth moment of the innovation process, see Equation (2.10). In Table 2.1, the critical values for d^3 are listed for "normal" innovations, i.e. $\mu_3 = 0$ and $\mu_4 = 3$. In the "non-normal" situation, the critical values have to be computed by simulation in each individual case. For the empirical investigations, we use the standardized residuals for estimating μ_3 and μ_4 of the innovation process.

Using the theoretical results presented in this section, we analyze log-returns of stock prices observed at the European market and the US market. The statistical analysis includes parameter estimation and testing for conditional homoscedasticity. Furthermore, the impact on applications in finance is discussed, where we focus on Value-at-Risk calculation.

2.3 Parameter Estimation and Testing

Choosing an appropriate model is an important and difficult task—not only for applications like Value-at-Risk calculation. In this section, we compare the models presented in Section 2 empirically. In particular, we examine log-returns of stock prices observed at the German market and the US market for conditional heteroscedasticity. We estimate the parameters of the models, and proceed by testing for reduction to the Black&Scholes model. We explicitly show the impact of the size of the alternative on the test result. The more alternatives are offered, the more likely is the rejection of the null hypothesis.

In the following, we analyze the daily log-returns of Allianz, BASF, Deutsche Telekom, VW, Apple, and IBM. The observed time period ranges from Sep. 1, 1996 to Sep. 1, 2000—with exception of Deutsche Telekom that was first listed Nov. 18, 1996. This includes 1044 data points for the entire 4 year horizon. In addition, we examine the most recent 2 years and the final year of the given time horizon, including 523 data points and 262 data points respectively. Detailed estimation results are reported in the Appendix. We focus on testing the null hypothesis of white noise, where the critical values are reported in Table 2.1.

The log-returns of Allianz and VW show strong evidence for conditional heteroscedasticity. We observe significant ARCH and GARCH effects, see Table 2.13 and Table 2.14 in the Appendix. The low standard errors indicate that the data fits into the time-series framework. The interpretation of the estimation results is validated by the test on model reduction to white noise versus various alternatives. For all investigated time horizons and both stocks—Allianz and VW, the null hypothesis of white noise is rejected for all admissible alternatives, see Table 2.2.

The test results of Deutsche Telekom and Apple are given in Table 2.3. Deutsche Telekom shows conditional heteroscedasticity. For the 1-year and 2-year horizon, we estimate a low ARCH effect that is not even very significant. However, the GARCH coefficient appears to be important, since the likelihood clearly improves for the model enhanced by the GARCH

Test results (Allianz)	$\widehat{d^1}$	$\widehat{d^2}$	$\widehat{d^3}$
1996 - 2000	15.24***	24.37***	97.61***
1998 - 2000	6.76***	9.20***	42.53***
1999 - 2000	7.52***	7.64***	50.25***
Test results (VW)	$\widehat{d^1}$	$\widehat{d^2}$	$\widehat{d^3}$
1996 - 2000	16.43***	43.16***	87.57***
1998 - 2000	13.45***	32.71***	64.88***
1999 - 2000	8.77***	8.77***	15.90***

Table 2.2: Deviance and rescaled deviance statistics for Allianz and VW. */**/*** denotes rejection of the null hypothesis for the significance level of 90/95/99%.

parameter δ, see Table 2.15 in the Appendix. Accordingly, we expect the result of the test for reduction to white noise to depend significantly on the set of given alternatives. The ARCH(1) alternative is not matching the conditional heteroscedasticity effects of Deutsche Telekom, hence the null hypothesis of white noise is accepted even for the 90% significance level. The GARCH(1,1) model provides the more appropriate set of alternatives. Here, the null hypothesis of white noise is clearly rejected in all cases. The same holds of course for the AR(1)-GARCH(1,1) alternative.

Test results (D. Telekom)	$\widehat{d^1}$	$\widehat{d^2}$	$\widehat{d^3}$
1996 - 2000	8.71***	86.15***	177.35***
1998 - 2000	0.67	19.17***	31.81***
1999 - 2000	0.11	7.82***	12.56***
Test results (Apple)	$\widehat{d^1}$	$\widehat{d^2}$	$\widehat{d^3}$
1996 - 2000	4.87**	12.41***	49.33***
1998 - 2000	4.48**	5.54**	13.65***
1999 - 2000	1.33	1.72	6.85**

Table 2.3: Deviance and rescaled deviance statistics for Deutsche Telekom and Apple. */**/*** denotes rejection of the null hypothesis for the significance level of 90/95/99%.

For Apple, we also notice that the acceptance or rejection of the white noise null hypothesis is influenced by the set of alternatives. The ARCH parameter λ is slightly significant, whereas δ is estimated with a remarkable high standard error, especially for the 1-year and 2-year horizon, see Table 2.15. Thus, the hypothesis of white noise tends to be rejected when the set of alternatives captures autoregression. This fact becomes apparent particularly for the 1-year horizon, where the likelihood increases substantially when introducing the autoregression parameter.

Finally, we observe the BASF and IBM data, see Table 2.4. The null hypothesis of white noise cannot be rejected for almost all time horizons and significance levels. Thus BASF and IBM are standard examples for log-returns of Black&Scholes type. For this kind of data, parameter estimation becomes complicated, since the information matrix is asymptotically singular for the presented models incorporating GARCH effects, i.e. $\delta \geq 0$, see

Test results (BASF)	$\widehat{d^1}$	$\widehat{d^2}$	$\widehat{d^3}$
1996 - 2000	10.08***	17.08***	27.32***
1998 - 2000	0.32	5.28**	7.32**
1999 - 2000	0.00	0.13	1.29
Test results (IBM)	$\widehat{d^1}$	$\widehat{d^2}$	$\widehat{d^3}$
1996 - 2000	2.00*	4.46**	15.58**
1998 - 2000	0.00	0.04	1.05
1999 - 2000	0.00	0.03	0.17

Table 2.4: Deviance and rescaled deviance statistics for BASF and IBM. */**/*** denotes rejection of the null hypothesis for the significance level of 90/95/99%.

the discussion in Andrews (1999b). If the data is white noise, we have to apply the procedure proposed in Section 2. For fixed δ, we maximize the likelihood function, and this is carried out for $\delta \in \Delta$, where we of course choose a finite set, i.e. $\delta \in \{0, 0.01, ..., 0.95\}$. We take the supremum of the maximized likelihood function depending on δ and compute the deviance statistics. In this case, the parameter δ is reported with no standard error of course, since it is more a nuisance parameter than an estimate, see Table 2.17 and Table 2.18. Nevertheless, we are able to run the maximum likelihood estimation procedure for some data close to *iid*, despite of the theoretical and also numerical problems that result from an (almost) singular information matrix, e.g., Apple, see Table 2.16.

Dealing with "white noise" data, the numerical procedure often overextends the estimation tools for GARCH of standard software packages. Brooks et al. (2001) discuss the accuracy of GARCH(1,1) model estimation in a well-conditioned setting. Here, we compute the maximum likelihood by a Newton-Raphson scheme, where we use the analytic gradient and Hessian matrix, what is close to the benchmark given by Brooks et al. (2001) in the sense of estimation accuracy.

The discussed LR tests are appropriate methods for model choice, but the computation of the asymptotic distribution of the deviance may become challenging, what was shown in Section 2. Besides, there exist other (weaker) criteria for selecting a model in the "best" way. Akaike's Information Criterion (AIC) is the most commonly used and is given by

$$\text{AIC} = -2\,\mathcal{L}(\hat{\theta}) + 2\,p\,, \tag{2.27}$$

where $\mathcal{L}(\hat{\theta})$ is the maximized log likelihood function and p denotes the number of parameters, see Chatfield (2001). We cross-check the LR test results with respect to the AIC, see Table 2.12. AIC prefers the Black&Scholes model exactly, when the LR-test results accepts the null hypothesis of white noise on the 10% level. In all other cases AIC suggests to choose the alternative time-series model subject to the LR test.

In the following section, we employ the results carried out here. Especially, we study the impact of model choice, of course within the presented framework, on Value-at-Risk calculation, where we are not only concerned with the VaR quality in terms of prediction accuracy, but also tackle the issue of computability of the estimates.

2.4 VaR under Different Model Specifications

The focus of this application is to illustrate the test results of the previous section by studying the performance of the different models with respect to Value-at-Risk calculation. Here, VaR calculation is a one-day prediction of a conditional quantile for a fixed level γ. Generally, the quality of the VaR calculation can be measured by standard backtesting according to Basle, see Jorion (2001). By this, we are enabled to evaluate each model due to the backtesting result. On the other hand, the likelihood ratio test results in Section 2.3 indicate which model to choose for fitting the data most adequately. In this section, the task is to compare the results of the backtesting procedure and the likelihood ratio test.

For each log-return series analyzed in Section 2.3, we perform a standard backtest. We use a 500-day history to estimate the parameter of each specific model in order to calculate the one-day VaR prediction on the 99% level. For the time-series, we now assume normal distributed innovations, hence the γ-VaR is given by

$$\mathrm{VaR}(\gamma) = -\mu_{t+1} + \sigma_{t+1}\Phi^{-1}(\gamma)\,, \tag{2.28}$$

where Φ is the standard normal distribution function, and μ_{t+1} is the mean value prediction, and σ_{t+1} the standard deviation prediction both based on the preceding 500 observations $X_t, ..., X_{t-499}$, and $\gamma = 99\%$. This is carried out for the last 500 days within the sample period, and for that period, we count the number of VaR exceptions. The Basle traffic light evaluates the backtesting result, i.e. the number of exceptions, by assigning "Green", "Yellow" or "Red". For the 99% level, the Green Zone ranges from 0 to 8, the Yellow Zone from 9 to 14, and the Red Zone starts with 15.

	AR-GARCH	GARCH	ARCH	Black&Scholes
Allianz	4 (G) / 5.637%	4 (G) / 5.663%	5 (G) 5.677%	5 (G) / 5.706%
VW	5 (G) / 5.395%	6 (G) / 5.434%	4 (G) 5.888%	3 (G) / 6.037%
D. Telekom	5 (G) / 7.309%	5 (G) / 7.316%	11 (Y) 6.471%	12 (Y) / 6.393%
Apple	10 (Y) / 8.739%	8 (G) / 8.760%	8 (G) 8.795%	8 (G) / 8.774%
BASF	5 (G) / 4.685%	4 (G) / 4.690%	4 (G) 4.681%	4 (G) / 4.685%
IBM	7 (G) / 5.387%	8 (G) / 5.392%	6 (G) 5.255%	6 (G) / 5.255%

Table 2.5: Backtesting results, i.e. number of exceptions including traffic light according to Basle (Green,Yellow,Red) and average Value-at-Risk.

The backtesting result is given in Table 2.5. With respect to VaR calculation, the number of exceptions together with the Basle traffic light characterize the quality of the model from the regulator's point of view. As well, we report the average VaR. A competing interest of financial institutions is to minimize the VaR as much as possible, since they have to keep a certain amount of their own capital proportional to the VaR. Roughly speaking, we examine each model for its risk in the sense of Basle and for its cost, where we interpret cost as own capital requirement.

For data with non negligible conditional heteroscedasticity effects, i.e. Allianz, VW, Deutsche Telekom, the backtesting results suggest to choose the more complex models like AR(1)-GARCH(1,1) and GARCH(1,1). In the case of VW, the number of exceptions is equal to 5 for all models, but the price in form of the average VaR increases considerably for the more simple models, e.g. the average VaR of the Black&Scholes model exceeds the average VaR of AR(1)-GARCH(1,1) by 12%. The Black&Scholes model and the ARCH(1) model have a significantly lower average VaR for Deutsche Telekom, however they also exhibits a clear "Yellow" traffic light with 11 and 12 exceptions.

Reviewing the test results for Apple, BASF, and IBM, the data that is close to white noise, the more simple Black&Scholes and the ARCH(1) model should be chosen. The average VaR attains for all models approximately the same value for each stock, but the number of exceptions tends to increase for the more complex models. The larger number of exceptions for the models incorporating the GARCH-component arises primarily from the numerical problems within the estimation procedure. For data close to white noise, the information matrix may become singular and consequently, the MLE is not reliable. The estimation procedure occasionally creates artificial and misleading effects that result in poor VaR predictions. For data close to white noise, the more complex models involving a GARCH-component are not advisable.

2.5 Studying the Power Function

In the preceeding section the test results for testing on reduction to white noise in a conditional heteroscedastic setting are illustrated by studying the impact on Value-at-Risk calculation. We found, that the LR test selected in majority of cases the model that exhibits the most suitable backtesting result, for $\gamma = 99\%$, and $T = 500$, cf. Tables 2.2-2.4 and Table 2.5. In the following, we examine the quality of the proposed LR test in a more statistical fashion by analyzing the power function. We conduct a simulation study where we focus on the ARCH(1) and the AR(1)-ARCH(1) model specification. Especially, we emphasize the effect of heavy tailed innovations on the power function and furthermore, we analyze the sample size of 500 that is used to calculate a Value-at-Risk.

The critical functions ϕ of the proposed tests take a quite canonical form, i.e. for a given significance level γ, the critical value is the quantile of the deviance statistics k_γ given in Table 2.1, and $\phi(x) = \mathbf{1}_{\{d(x) > k_\gamma\}}$, where d denotes the deviance. The power function is given by

$$\beta(\theta) = \mathbb{E}_\theta\{\phi(X)\}\,, \qquad (2.29)$$

where θ parameterizes the model, and X is the random experiment/time series.

A main concern of this section is to evaluate the power function of the proposed LR test, For this purpose, we partition the parameter space Ω into three disjoint sets

$$\Omega = \Omega_H \cup \Omega_I \cup \Omega_K,$$

of which Ω_K is a subset of the alternative such that $\inf_{\theta \in \Omega_K} \beta(\theta) \geq \gamma$, and Ω_I designates the *indifference zone*, e.g., see Lehmann (1986), Ch. 9. From the definition of the indifference zone Ω_I it is clear that Ω_I collects the parameters θ of the alternative such that

$$\beta(\theta) < \gamma\,, \quad \text{for all } \theta \in \Omega_I\,,$$

and hence on Ω_K the error of second kind (accepting the null hypothesis when it is false) is bounded by $1-\gamma$. Recall that by construction of the test, the error of first kind (rejecting the null hypothesis when it is true) equals $1-\gamma$, i.e. $\beta(\theta) = 1-\gamma$, for $\theta \in \Omega_H$. In the following, we will emphasize the indifference zone. On this set statistical decisions are a delicate issue, since the LR test shows a lack of power there.

It is important to note that the AR(1)-ARCH(1) time series X parameterized by $\theta = (\alpha, \omega, \lambda)$ can be rescaled. The time series given by $X/\sqrt{\omega}$ is described by $(\alpha, 1, \lambda)$. Taking into account this relation, it is sufficient to study the time series with restricted parameterization $\omega = 1$. Of course, this argumentation applies also to the ARCH(1) specification.

2.5.1 ARCH(1) Model

The ARCH(1) model is given by $\theta = (1, \lambda)$, after applying the proper scaling scheme discussed above, and w.l.o.g. we can fix ω and $\Omega = [0, 1)$, where $\lambda < 1$ for stationarity reasons. We evaluate the power function $\beta(\lambda)$ on the 95%-level for 3 sample sizes $T = 100, 500, 2500$, where we first emphasize Gaussian innovations. The critical values are calculated/simulated explicitly for the considered sample sizes by performing 10.000 Monte Carlo runs for each critical value. The results are reported in Table 2.6.

sample size T	100	500	2500	∞
critical value $k_{95\%}$	2.17	2.52	2.63	2.71

Table 2.6: Critical values for testing the white noise model vs. the ARCH(1) alternative on the 95% level for Gaussian innovations.

A simulation study is conducted for the power function $\beta(\lambda)$. For the sample sizes $T = 100, 500, 2500$, we perform 10.000 Monte Carlo runs for computing $\beta(\lambda_i)$ for discrete values $0 = \lambda_0 < ... < \lambda_n < 1$. Figure 2.1 shows the power function for the LR test on white noise for ARCH(1) alternative and Gaussian innovations. The dotted line represents the γ-level, where $\gamma = 95\%$ is the significance level of the test. The intersection of this line with the power function gives us the minimal λ_I such that the error of second kind is bounded below α, thus $\Omega_I =]0, \lambda_I[$.

For $T = 100$ the power of the LR test is quite weak, and it seems hardly possible to establish empirical results with statistical significance for this sample size. The minimal heteroscedasticity parameter to bound the error of second kind from above by $1-\gamma = 5\%$ takes the value $\lambda_I = 0.58$, and $\Omega_I =]0, 0.58[$. The power improves when the sample size T is increased from 100 to 500. Here, $\lambda_I = 0.19$ implying an indifference zone $\Omega_I =]0, 0.19[$, and for moderate heteroscedasticity, e.g., $\lambda \approx 0.2, ..., 0.3$, the test provides a satisfying power well above 95%. And for $T = 2500$, we find $\lambda_I = 0.07$ and the indifference zone becomes relatively narrow $\Omega_I =]0, 0.07[$, indicating the claimed power even for rather little heteroscedasticity effects.

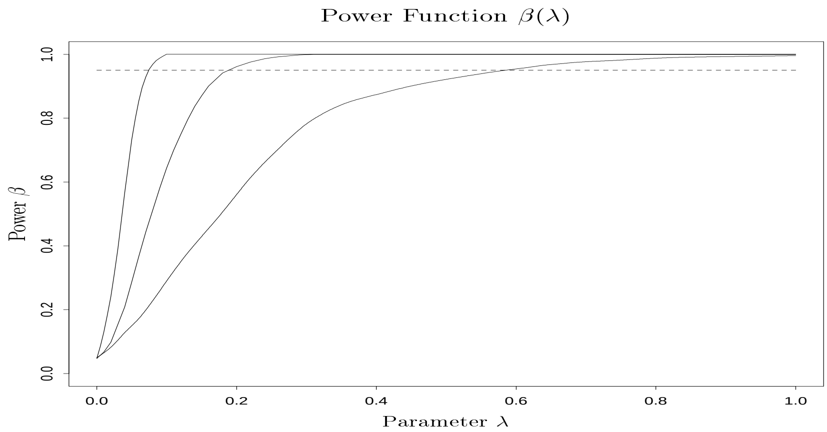

Figure 2.1: Power function $\beta(\lambda)$ for testing on white noise with ARCH(1) alternative, for Gaussian innovations, and sample sizes $T = 100, 500, 2500$.

sample size T	100	500	2500	∞
critical value $k_{95\%}$	2.19	2.31	2.73	2.71

Table 2.7: Critical values for testing the white noise model vs. the ARCH(1) alternative on the 95% level for t_5-distributed innovations.

Theorem 2.3 remains true if we move further, from Gaussian white noise to more general innovation types. The sole restriction within Theorem 2.3 is that the fourth moment μ_4 of the innovations is assumed to exist. For applications, especially in finance, the tail behaviour of the innovations plays a central role. In the following, we highlight the issue of heavy tailed innovations by applying the t_5-distribution for the innovations. It is well known for the t_5-distribution that the fourth moment exists $\mu_4 = 9$, and the distribution function is heavy tailed with tail index $\xi = 0.2$, what is quite common for financial data.

The critical values for finite sample sizes $T = 100, 500, 2500$, and t_5-distributed white noise as innovation process are reported in Table 2.7. Based on this critical values the power $\beta(\lambda)$ of the specific LR test is computed, again by 10.000 Monte Carlo runs for each λ_i, $0 = \lambda_0 < ... < \lambda_n < 1$.

The heavy tails of the t_5-innovations reduce the power of the LR test, see Figure 2.2. This effect becomes very much apparent when considering the sample size $T = 100$. Here, the power function $\beta(\lambda)$ does not attain the 95%-level even for the largest parameter $\lambda = 1$, thus $\Omega_I =]0, 1[$ and $\Omega_K = \emptyset$. In this case, the power of the LR test on the alternative is such low that we can not control the error of second kind properly, and hence we accept the null hypothesis of white noise though it is false with a relatively high probability. For the sample size $T = 500$, suitable power on the alternative is attained for $\lambda \geq \lambda_I = 0.28$, and we realize a deterioration compared to the Gaussian case where $\lambda_I = 0.19$. Only for the sample size $T = 2500$ the loss of power is not that noticeable, for the t_5-innovations we observe $\lambda_I = 0.09$ and $\Omega_I =]0, 0.09[$, in relation to $\lambda_I = 0.07$ in the Gaussian setting. Table 2.8 summarizes the comparison of the indifference zone subject to varying innovation types.

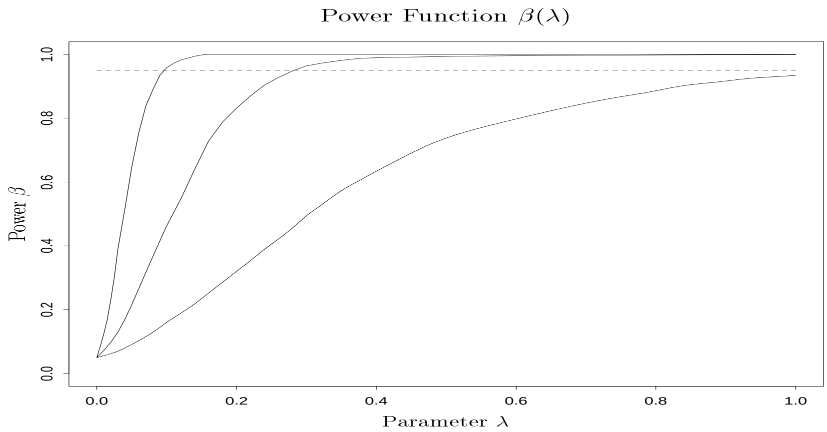

Figure 2.2: Power function $\beta(\lambda)$ for testing on white noise with ARCH(1) alternative, for t_5-distributed innovations, and sample sizes $T = 100, 500, 2500$.

sample size T	100	500	2500
Gaussian innovations	$]0, 0.58[$	$]0, 0.19[$	$]0, 0.07[$
t_5 innovations	$]0, 1.00[$	$]0, 0.28[$	$]0, 0.09[$

Table 2.8: The indifference zone Ω_I for testing the white noise model vs. the ARCH(1) alternative on the 95% level for Gaussian and t_5-distributed innovations.

2.5.2 AR(1)-ARCH(1) Model

For the AR(1)-ARCH(1) model, we apply the restricted parameterization discussed above by fixing ω and emphasizing autoregression and ARCH-effects $\theta = (\alpha, \lambda)$ for Gaussian innovations. Thus, the parameter space is given by $\Omega =]-1, 1[\times [0, 1[$. Table 2.9 reports the critical values for LR testing on white noise within an AR(1)-ARCH(1) model for finite sample sizes, where all displayed values are computed by using 5.000 Monte Carlo runs.

sample size T	100	500	2500	∞
critical value $k_{95\%}$	4.03	4.64	4.68	5.13

Table 2.9: Critical values for testing the white noise model vs. the AR(1)-ARCH(1) alternative on the 95% level for Gaussian innovations.

The power function $\beta(\theta)$ is a function $]-1, 1[\times [0, 1[\to [0, 1]$. Since we need to simulate a finite set of points in two dimensions, we reduce the illustration of the power function by presenting the indifference zone for the sample sizes $T = 100, 500, 2500$, see Figure 2.3. The upper line represents the boundary for $T = 100$, and below the indifference zone Ω_I is situated except for the origin $\Omega_H = \{(0, 0)\}$. The ARCH parameter λ primarily determines the behaviour of the power function/indifference zone, for small autoregression λ must exceed 0.70 in order to provide the claimed power of the test of 95%. For the sample size $T = 100$ testing the null hypothesis of white noise is hardly possible, since the error of second kind can rather be controlled in the area of interest for empirical applications. The power improves, the indifference zone becomes significantly smaller, when the sample size increases to $T = 500$. And this effect continues when we have $T = 2500$. In the latter case, the indifference zone can be considered as a rather small region around the origin representing the null hypothesis $\Omega_H = \{(0,0)\}$. For the AR(1)-ARCH(1) model we do not discuss heavy tailed innovations, since the calculation of the indifference zone is a computationally demanding task.

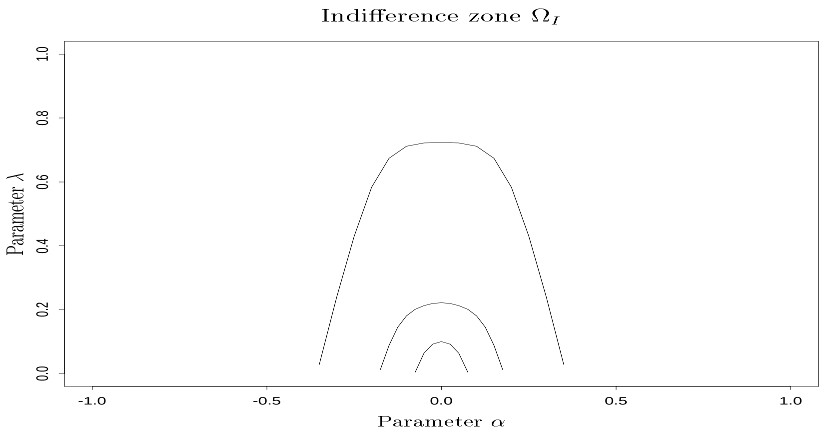

Figure 2.3: The indifference zone Ω_I for testing on white noise with AR(1)-ARCH(1) alternative, for Gaussian innovations, and sample sizes $T = 100, 500, 2500$.

In this section, the sample size effect on the power function of the LR test for testing on reduction to white noise is investigated by a simulation study. In the following, the results are applied to a practical issue: Value-at-Risk (VaR) calculation, where especially the sample size $T = 500$ becomes important, since this is a possible time horizon/sample size for the VaR application.

2.5.3 Empirical Investigations

In this section we study the impact of the above results concerning the power function of the LR test for reduction to white noise within an ARCH(1) setting, respectively AR(1)-ARCH(1). We aim at VaR calculation as application in finance, and hence use a sample size/time horizon of $T = 500$. Besides $T = 250$, this is a standard sample size, see Jorion (2001). We investigate the log-returns of stock prices and indices from the German market and the US market, in particular, Apple, BASF, IBM, Volkswagen (VW), the DAX index, and the NASDAQ index. We use data from 1.1.1994 to 31.12.1995 what is a different time period compared to Section 2.3. For this study, we assume the innovations to be Gaussian affecting the LR test w.r.t. the critical value and the power function.

The empirical results for the ARCH(1) model are reported in Table 2.10, including the maximum likelihood estimate of the ARCH parameter λ, the deviance statistics d, and the power $\beta(\lambda)$ evaluated at the point estimate. Applying the LR test, the null hypothesis of white noise is accepted for IBM and DAX at the 95% level. In the other four cases the null is rejected, but only for Apple the point estimate of λ is not inside the indifference zone

	$\hat{\lambda}$	d	$\beta(\hat{\lambda})$
Apple	0.1862	15.82*	0.95
BASF	0.1346	5.77*	0.81
IBM	0.0102	0.14	0.07
VW	0.1574	11.26*	0.89
DAX	0.0280	0.59	0.13
NASDAQ	0.1518	13.84*	0.87

Table 2.10: Results of the LR test on reduction to white noise with ARCH(1) alternative, including estimate of the ARCH parameter λ, the deviance, and the power $\beta(\hat{\lambda})$ evaluated at the point estimate. * denotes rejection of the null at a significance level of 95% .

$\Omega_I =]0, 0.19[$, see Table 2.8, implying a power function evaluated at the point estimate above 95%. For BASF, VW, and NASDAQ, the null is rejected with an "estimated" power of 81%, 89%, and 87%.

The same program as above is now considered for the model extended by an autoregressive component of order one, see Table 2.11. Note first, a rejection of the null hypothesis of white noise is more likely for this specification, because the alternative can also capture autoregressive structures. The null hypothesis is still not rejected for IBM and DAX. For IBM we observe a noticeable autoregression with $\hat{\alpha} = -0.0699$ and a corresponding t-value of 1.57. This increases the deviance from 0.14 to 2.33, though we are still below the critical value of 4.63, see Table 2.9. For the NASDAQ time series we note a similar behaviour: $\hat{\alpha} = 0.2458$ and a corresponding t-value of 4.70, and the deviance increases from 13.84 to 35.80.

	$\hat{\alpha}$	$\hat{\lambda}$	d	$\beta(\widehat{(\alpha, \lambda)})$
Apple	0.0574	0.1952	17.16*	0.94
BASF	-0.0415	0.1420	6.49*	0.80
IBM	-0.0699	0.0000	2.33	0.27
VW	0.0848	0.1453	14.04*	0.89
DAX	-0.0081	0.0282	0.62	0.11
NASDAQ	0.2458	0.1846	35.80*	1.00

Table 2.11: Results of the LR test on reduction to white noise with AR(1)-ARCH(1) alternative, including estimate of the autoregression α and ARCH parameter λ, the deviance, and the power $\beta(\widehat{(\alpha, \lambda)})$ evaluated at the point estimate. * denotes rejection of the null at a significance level of 95% .

The maximum likelihood estimates of the autoregressive component α and the ARCH parameter λ, and their impact on the power function is illustrated in Figure 2.4. The indifference zone Ω_I is given by the area below the curve excluding the origin representing the null Ω_H. Considering each parameter separately, $|\alpha| \geq 0.18$ or $\lambda \geq 2.20$ is a sufficient condition to guarantee a minimal power $\beta(\theta)$ of 95%. For the investigated data, the rejected null hypothesis is indicated by a dot, and a triangle gives evidence that the null is not rejected. Within the sample, only NASDAQ rejects the null hypothesis with an "estimated" power $\beta(\hat{\theta}) = 100\%$ exceeding the demanded level of 95%.

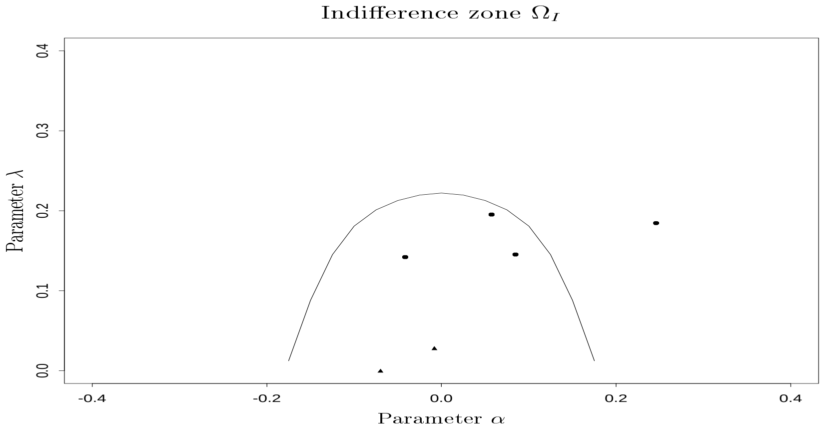

Figure 2.4: The indifference zone Ω_I for testing on white noise with AR(1)-ARCH(1) alternative, for Gaussian innovations, and sample sizes $T = 500$, including the point estimates for IBM, BASF, DAX, Apple, VW, NASDAQ (from left to right).

The choice of an appropriate model for describing financial time series is an important issue, especially when VaR calculation is considered, see e.g. Christoffersen et al. (2001) for discussion. In this section, we analyze stereotype models allowing for conditional heteroscedasticity and autoregression. We find that a decision inbetween the proposed models on the basis of statistical concepts is hardly possible, even for the "large" sample size $T = 500$. The LR test does not reject the null of white noise in two of six cases, and when rejecting, the "estimated" power of the test falls below the required power. The empirical results of the tests are heterogeneous and lack statistical power. Graphically speaking, the area of the indifference zone is just too large for the considered setting.

2.6 Conclusion

In this chapter we compare financial time-series models allowing for conditional heteroscedasticity and autoregression. Primarily, we utilize the likelihood ratio test for the comparison of the different models, and cross-check the LR result by applying the AIC concept, and also, we perform standard backtests according to Basle. In general, we can not find evidence for preferring a specific model for all observed log-returns. We are suggested to use the more simple Black&Scholes model for data close to white noise, and we ought to rely on GARCH-type models, whenever the data exhibits conditional heteroscedasticity. This result is striking, especially in the case of backtesting, since the largest model, the AR(1)-GARCH(1,1) model, incorporates all other models discussed here. And hence, we would not expect heterogeneous results, since the largest model should cover all possible effects. The reason for this can be found in the numerics of the estimation, i.e. the information matrix becomes singular when we apply GARCH models to white noise data. We believe that it is not possible to find a "benchmark model" for describing financial time-series, and the problem of model choice has to be discussed in each specific case, also depending on the application.

Additionally, we examine the quality of the proposed LR test in a more statistical manner by analyzing the power function. We conduct a simulation study where we focus on the ARCH(1) and the AR(1)-ARCH(1) model specification. For the ARCH(1) model, also the effect of heavy tailed innovations is investigated, and the findings strongly indicate that the power function deteriorates significantly when dealing with heavy tails. For the sample size $T = 500$ that is often applied in finance, the indifference zone of the LR tests is of particular interest. The data analyzed here, exhibits mostly conditional heteroscedasticity, but these effects are not strong enough to reject the null of white noise with the power required, since the maximum likelihood estimates are situated inside the indifference zone of the considered tests. For the considered time horizon $T = 500$, statistical methods are barely a sufficient basis for selecting an appropriate model. Generally, the model choice depends on the intended application and hence, the quality of this model should be measured by backtesting the performance of the model empirically.

2.7 Tables: AIC, and Estimation Results

AIC for Allianz	Black&Scholes	ARCH	GARCH	AR-GARCH
1996 - 2000	-4922.32	-4980.38	-5014.34	-5013.94
1998 - 2000	-2363.00	-2391.78	-2400.86	-2399.54
1999 - 2000	-1197.80	-1241.98	-1240.70	-1242.04
AIC for VW	Black&Scholes	ARCH	GARCH	AR-GARCH
1996 - 2000	-4898.76	-4937.00	-4978.46	-4980.32
1998 - 2000	-2422.44	-2444.92	-2477.96	-2481.32
1999 - 2000	-1318.24	-1331.86	-1329.86	-1328.14
AIC for D. Telekom	Black&Scholes	ARCH	GARCH	AR-GARCH
1996 - 2000	-4315.64	-4331.22	-4485.32	-4487.00
1998 - 2000	-2092.38	-2091.40	-2117.48	-2118.18
1999 - 2000	-1018.66	-1016.82	-1025.44	-1025.24
AIC for Apple	Black&Scholes	ARCH	GARCH	AR-GARCH
1996 - 2000	-3873.46	-3889.60	-3915.64	- 3916.80
1998 - 2000	-1909.60	-1913.28	-1912.64	-1917.24
1999 - 2000	-913.86	-913.38	-911.82	-914.72
AIC for BASF	Black&Scholes	ARCH	GARCH	AR-GARCH
1996 - 2000	-5267.96	-5282.04	-5287.22	-5289.28
1998 - 2000	-2597.50	-2595.94	-2600.80	-2598.80
1999 - 2000	-1308.96	-1306.96	-1305.10	-1304.10
AIC for IBM	Black&Scholes	ARCH	GARCH	AR-GARCH
1996 - 2000	-4920.84	-4925.62	-4931.94	-4930.42
1998 - 2000	-2350.28	-2348.28	-2346.40	-2345.32
1999 - 2000	-1130.74	-1128.74	-1126.84	-1124.90

Table 2.12: AIC for Allianz, VW, Deutsche Telekom, Apple, BASF, and IBM.

Sep. 1., 1996 - Sep. 1., 2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.5213	-	-	2462.16
s.e.	(-)	(0.0230)	(-)	(-)	
ARCH(1)	-	0.4125	0.2101	-	2492.19
s.e.	(-)	(0.0220)	(0.0449)	(-)	
GARCH(1,1)	-	0.0740	0.1630	0.7062	2510.17
s.e.	(-)	(0.0098)	(0.0340)	(0.0073)	
AR(1)-GARCH(1,1)	0.0456	0.0714	0.1629	0.7113	2510.97
s.e.	(0.0365)	(0.0110)	(0.0374)	(0.0196)	
Sep. 1., 1998 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.6308	-	-	1182.50
s.e.	(-)	(0.0394)	(-)	(-)	
ARCH(1)	-	0.4863	0.1379	-	1197.89
s.e.	(-)	(0.0390)	(0.0690)	(-)	
GARCH(1,1)	-	0.1961	0.2359	0.4779	1203.43
s.e.	(-)	(0.0624)	(0.0782)	(0.1303)	
AR(1)-GARCH(1,1)	0.0493	0.1747	0.2240	0.5213	1203.77
s.e.	(0.0636)	(0.0620)	(0.0666)	(0.1193)	
Sep. 1., 1999 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.5904	-	-	599.90
s.e.	(-)	(0.0511)	(-)	(-)	
ARCH(1)	-	0.3112	0.6509	-	622.99
s.e.	(-)	(0.0414)	(0.1755)	(-)	
GARCH(1,1)	-	0.2865	0.6509	0.0430	623.35
s.e.	(-)	(0.0589)	(0.1559)	(0.0608)	
AR(1)-GARCH(1,1)	-0.1253	0.2823	0.7377	0.0130	625.02
s.e.	(0.0630)	(0.0428)	(0.1790)	(0.0350)	

Table 2.13: Parameter estimates with standard errors (s.e.) in parentheses, and the maximum log likelihood (ML) for Allianz.

Sep. 1., 1996 - Sep. 1., 2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.5356	-	-	2450.38
s.e.	(-)	(0.0234)	(-)	(-)	
ARCH(1)	-	0.4412	0.1761	-	2470.50
s.e.	(-)	(0.0138)	(0.0320)	(-)	
GARCH(1,1)	-	0.0178	0.0665	0.9024	2492.23
s.e.	(-)	(0.0053)	(0.0120)	(0.0178)	
AR(1)-GARCH(1,1)	0.0654	0.0180	0.0664	0.9020	2494.16
s.e.	(0.0314)	(0.0054)	(0.0125)	(0.0184)	
Sep. 1., 1998 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.5679	-	-	1212.22
s.e.	(-)	(0.0351)	(-)	(-)	
ARCH(1)	-	0.4391	0.2348	-	1224.46
s.e.	(-)	(0.0283)	(0.0562)	(-)	
GARCH(1,1)	-	0.0136	0.0577	0.9157	1241.98
s.e.	(-)	(0.0067)	(0.0142)	(0.0230)	
AR(1)-GARCH(1,1)	0.1022	0.0115	0.0522	0.9246	1244.66
s.e.	(0.0421)	(0.0060)	(0.0136)	(0.0211)	
Sep. 1., 1999 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.3721	-	-	660.12
s.e.	(-)	(0.0326)	(-)	(-)	
ARCH(1)	-	0.2881	0.2132	-	667.93
s.e.	(-)	(0.0321)	(0.0835)	(-)	
GARCH(1,1)	-	0.2881	0.2132	0.0000	667.93
s.e.	(-)	(0.0351)	(0.0841)	(0.0481)	
AR(1)-GARCH(1,1)	0.1083	0.2892	0.2045	0.0000	668.07
s.e.	(0.0708)	(0.0499)	(0.0849)	(0.1238)	

Table 2.14: Parameter estimates with standard errors (s.e.) in parentheses, and the maximum log likelihood (ML) for VW.

Sep. 1., 1996 - Sep. 1., 2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.7439	-	-	2158.82
s.e.	(-)	(0.0335)	(-)	(-)	
ARCH(1)	-	0.6417	0.1446	-	2167.61
s.e.	(-)	(0.0230)	(0.0380)	(-)	
GARCH(1,1)	-	0.0067	0.0678	0.9264	2245.66
s.e.	(-)	(0.0036)	(0.0111)	(0.0104)	
AR(1)-GARCH(1,1)	0.0652	0.0066	0.0702	0.9244	2247.50
s.e.	(0.0361)	(0.0021)	(0.0115)	(0.0106)	
Sep. 1., 1998 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	1.0675	-	-	1047.19
s.e.	(-)	(0.0660)	(-)	(-)	
ARCH(1)	-	1.0175	0.0480	-	1047.70
s.e.	(-)	(0.0585)	(0.0466)	(-)	
GARCH(1,1)	-	0.0116	0.0362	0.9530	1061.74
s.e.	(-)	(0.0072)	(0.0114)	(0.0135)	
AR(1)-GARCH(1,1)	0.0739	0.0115	0.0370	0.9522	1063.09
s.e.	(0.0488)	(0.0071)	(0.0115)	(0.0135)	
Sep. 1., 1999 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	1.1903	-	-	510.33
s.e.	(-)	(0.1040)	(-)	(-)	
ARCH(1)	-	1.1583	0.0281	-	510.41
s.e.	(-)	(0.1039)	(0.0609)	(-)	
GARCH(1,1)	-	0.2300	0.1328	0.6766	515.72
s.e.	(-)	(0.1670)	(0.0700)	(0.2002)	
AR(1)-GARCH(1,1)	0.0901	0.2049	0.1220	0.7064	516.62
s.e.	(0.0753)	(0.1607)	(0.0659)	(0.1931)	

Table 2.15: Parameter estimates with standard errors (s.e.) in parentheses, and the maximum log likelihood (ML) for Deutsche Telekom.

Sep. 1., 1996 - Sep. 1., 2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	1.4301	-	-	1937.73
s.e.	(-)	(0.0625)	(-)	(-)	
ARCH(1)	-	1.2671	0.1215	-	1946.80
s.e.	(-)	(0.0299)	(0.0368)	(-)	
GARCH(1,1)	-	0.4133	0.1175	0.5943	1960.82
s.e.	(-)	(0.0994)	(0.0343)	(0.0918)	
AR(1)-GARCH(1,1)	-0.0635	0.4150	0.1188	0.5911	1962.40
s.e.	(0.0365)	(0.1003)	(0.0344)	(0.0932)	
Sep. 1., 1998 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	1.5114	-	-	955.80
s.e.	(-)	(0.0936)	(-)	(-)	
ARCH(1)	-	1.3133	0.1379	-	958.64
s.e.	(-)	(0.0957)	(0.0690)	(-)	
GARCH(1,1)	-	0.8283	0.1330	0.3247	959.32
s.e.	(-)	(0.3426)	(0.0657)	(0.2517)	
AR(1)-GARCH(1,1)	-0.1262	0.9665	0.1375	0.2214	962.62
s.e.	(0.0495)	(0.3183)	(0.0650)	(0.2294)	
Sep. 1., 1999 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	1.7757	-	-	457.93
s.e.	(-)	(0.1551)	(-)	(-)	
ARCH(1)	-	1.5909	0.1076	-	458.69
s.e.	(-)	(0.1719)	(0.0978)	(-)	
GARCH(1,1)	-	1.1339	0.1225	0.2450	458.91
s.e.	(-)	(0.6179)	(0.0987)	(0.3718)	
AR(1)-GARCH(1,1)	-0.1508	1.4602	0.0832	0.0782	461.36
s.e.	(0.0672)	(0.8076)	(0.0915)	(0.4603)	

Table 2.16: Parameter estimates with standard errors (s.e.) in parentheses, and the maximum log likelihood (ML) for Apple.

Sep. 1., 1996 - Sep. 1., 2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.3761	-	-	2634.98
s.e.	(-)	(0.0165)	(-)	(-)	
ARCH(1)	-	0.3345	0.1087	-	2643.02
s.e.	(-)	(0.0153)	(0.0312)	(-)	
GARCH(1,1)	-	0.0346	0.0602	0.8484	2646.61
s.e.	(-)	(0.0133)	(0.0162)	(0.0475)	
AR(1)-GARCH(1,1)	-0.0087	0.0342	0.0601	0.8491	2648.64
s.e.	(0.0325)	(0.0136)	(0.0164)	(0.0485)	
Sep. 1., 1998 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.4025	-	-	1299.75
s.e.	(-)	(0.0249)	(-)	(-)	
ARCH(1)	-	0.3886	0.0353	-	1299.97
s.e.	(-)	(0.0317)	(0.0559)	(-)	
GARCH(1,1)	-	0.0111	0.0001	0.9500	1303.40
s.e.	(-)	(0.0033)	(0.0082)	(fixed)	
AR(1)-GARCH(1,1)	0.0327	0.0114	0.0000	0.9500	1303.40
s.e.	(0.0445)	(0.0085)	(0.0119)	(fixed)	
Sep. 1., 1999 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.3857	-	-	655.48
s.e.	(-)	(0.0338)	(-)	(-)	
ARCH(1)	-	0.3857	0.0000	-	655.48
s.e.	(-)	(0.0478)	(0.0881)	(-)	
GARCH(1,1)	-	0.1870	0.0277	0.4900	655.55
s.e.	(-)	(0.0320)	(0.0751)	(fixed)	
AR(1)-GARCH(1,1)	0.0755	0.0198	0.0004	0.9500	656.10
s.e.	(0.0622)	(0.0285)	(0.0191)	(fixed)	

Table 2.17: Parameter estimates with standard errors (s.e.) in parentheses, and the maximum log likelihood (ML) for BASF.

Sep. 1., 1996 - Sep. 1., 2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.5244	-	-	2461.42
s.e.	(-)	(0.0230)	(-)	(-)	
ARCH(1)	-	0.4941	0.0589	-	2464.81
s.e.	(-)	(0.1245)	(0.0209)	(-)	
GARCH(1,1)	-	0.0249	0.0298	0.9237	2468.97
s.e.	(-)	(0.0083)	(0.0084)	(0.0204)	
AR(1)-GARCH(1,1)	-0.0251	0.0341	0.0294	0.9237	2469.21
s.e.	(0.0325)	(0.0085)	(0.0087)	(0.0210)	
Sep. 1., 1998 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.6464	-	-	1176.14
s.e.	(-)	(0.0400)	(-)	(-)	
ARCH(1)	-	0.6486	0.0000	-	1176.14
s.e.	(-)	(0.0420)	(0.0198)	(-)	
GARCH(1,1)	-	0.3746	0.0000	0.4300	1176.20
s.e.	(-)	(0.0261)	(0.0196)	(fixed)	
AR(1)-GARCH(1,1)	-0.0557	0.4696	0.0000	0.2800	1176.66
s.e.	(0.0444)	(0.0375)	(0.0195)	(fixed)	
Sep. 1., 1999 - Sep. 1.,2000	$\widehat{\alpha}$	$\widehat{\beta}$ $[10^{-3}]$	$\widehat{\lambda}$	$\widehat{\delta}$	ML
Black&Scholes	-	0.7633	-	-	566.37
s.e.	(-)	(0.0668)	(-)	(-)	
ARCH(1)	-	0.7655	0.0000	-	566.37
s.e.	(-)	(0.0704)	(0.0292)	(-)	
GARCH(1,1)	-	0.5022	0.0000	0.3600	566.42
s.e.	(-)	(0.0488)	(0.0317)	(fixed)	
AR(1)-GARCH(1,1)	-0.1173	0.4787	0.0000	0.3900	566.45
s.e.	(0.0632)	(0.0525)	(0.0314)	(fixed)	

Table 2.18: Parameter estimates with standard errors (s.e.) in parentheses, and the maximum log likelihood (ML) for IBM.

Chapter 3

Testing for Mean Reversion

3.1 Introduction

Stochastic processes of Ornstein-Uhlenbeck (OU) type are utilized for modeling purposes in many areas. They capture, in a continuous time setting, a possible *autoregressive* effect in the data. Besides engineering and physics, this class of processes is widely used in financial modelling. In applications, it is also often useful to include a mean level effect in the model. For example, Vasicek (1977) pioneered the application of such mean reverting stochastic processes for interest rate modeling. More recently, Barndorff-Nielsen and Shephard (2001) and Barndorff-Nielsen, Nicolato and Shephard (2002) (see also their references) studied non-Gaussian OU based models, especially stochastic volatility models, and presented some of their uses in financial economics. Schwartz (1997) applied the OU concept for modeling of commodity prices including derivative pricing and hedging.

The behaviour of an (ordinary) OU process changes dramatically when the autoregressive effect is not present. Then the process reduces to a Brownian motion. This results in a considerable simplification if it holds for the data we are analysing, since of course then the special features of Brownian motion and the normal distribution are available, especially, for example, for option pricing purposes. In this chapter, we test OU processes in continuous time for such a reduction, i.e., equivalently, for the absence of mean reversion effects.

Rather than restricting ourselves just to Brownian motion models, we consider continuous time OU models driven by Lévy process which are completely general other than being required to have finite variance. Thus, in particular, there may be jump components of the Lévy process present. There has been increasing interest in the use of Lévy processes in financial modelling, in particular, see, e.g., the volume by Barndorff-Nielsen, Mikosch and Resnick (2001).

For an appropriate discretized version of the model, we utilize likelihood ratio methods to test the null hypothesis of no autoregression. In the discrete time setting, this corresponds to a unit root test. As an additional complication, the mean reversion level turns out to be a nuisance parameter which cannot be identified under the null. Davies (1977, 1987) presents techniques to handle these kinds of problems. Using them, we are able to derive the asymptotic distribution of the pseudo-likelihood ratio statistic for testing this null. In the general Lévy process setting, the limit is taken for a refining sequence of partitions on a fixed time interval, where the mesh of the partitions tends to zero, and also as the length of the interval grows large. For the special case of Brownian Motion as driving process, we deduce the distribution in a quite explicit way, and find results related to and generalising the Dickey-Fuller theory.

3.2 The OU Model

In this section, we briefly recall the theory of Ornstein-Uhlenbeck processes driven by Lévy processes. We rely on Protter (1990) for results related to the OU stochastic differential equation, and refer to Sato (1999) and Bertoin (1996) for general background on Lévy processes.

On a filtered probability space $\left(\mathcal{S}, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0}\right)$ satisfying the "usual hypotheses" (see Protter, p. 3), we are given a càdlàg Lévy process $L = \{L_t : t \geq 0\}$. The distribution of L is completely described by its Lévy-Khintchine characteristics $(\alpha_L, \sigma_L^2, \nu_L)$, where α_L is the drift, σ_L^2 is the variance of the Brownian part of L, and ν_L is the Lévy measure. We assume the Lévy process has mean zero and finite second moment normalized to unity:

$$\mathbb{E}\{L_1\} = 0\,, \quad \int_{\mathbf{R}} x^2\,\nu_L(dx) < \infty\,, \quad \text{and} \quad \mathbb{E}\left\{L_1^2\right\} = 1.$$

Hence L is a square-integrable martingale, normalized so that $\mathrm{Var}(L_t) = \mathbb{E}\{L_t^2\} = t\,\mathbb{E}\{L_1^2\} = t$, for $t \geq 0$.

The process $X = \{X_t : t \geq 0\}$ with initial (fixed) value $X_0 \in \mathbb{R}$ is defined by the stochastic differential equation

$$dX_t = \gamma\,(m - X_{t-})\,dt + \sigma dL_t\,, \quad \text{for } t \geq 0, \tag{3.1}$$

where $\gamma \in \mathbb{R}$, $m \in \mathbb{R}$, and $\sigma > 0$. Equation (3.1) admits a unique solution, for whose existence see Protter (1990), Theorem 7, p. 197. Applying Itô's Formula, we can verify that

$$X_t = m + (X_0 - m)\,e^{-\gamma t} + \sigma \int_0^t e^{\gamma(s-t)}\,dL_s\,, \quad \text{for } t \geq 0. \tag{3.2}$$

Equation (3.2) highlights the underlying concept of the OU equation. Usually $\gamma > 0$ is assumed. Then, as $t \to \infty$, the deterministic part of X tends to the mean level m, where the perturbations caused by L are aggregated, but the influence of an infinitesimal perturbation ΔL decreases exponentially with the time passed since it occurred. When L is a standard Brownian Motion (SBM), X is a Gaussian process which fluctuates with volatility σ, and is pulled back to its mean level m at rate γ. For our purposes, however, we do not need to assume $\gamma > 0$, so we adopt the more general situation where $\gamma \in \mathbb{R}$.

The form of the solution changes when $\gamma = 0$. The process X is no longer mean reverting, but reduces to

$$X_t = X_0 + \sigma L_t \,, \quad \text{for } t \geq 0. \tag{3.3}$$

We find a pure Lévy process for X. The mean reversion level m disappears, since it vanishes from (3.1) and (3.2) when $\gamma = 0$. Since we wish to test precisely the hypothesis

$$\mathbf{H}_0 : \gamma = 0,$$

we are in the situation where the parameter m is said to "vanish under the null". Methods of dealing with this phenomenon in a likelihood setting are known, and we use them in the next section.

We also need to discretise. Continuous time models are useful for their theoretical properties, but in reality the trajectories of the process cannot be observed continuously, and the process must be sampled in discrete time. For estimation and testing purposes, a discrete time analogue of the continuous time model is required. For discretely observed diffusions, Sørensen (1997) reviews the concept of simple and martingale estimating functions. For more general Lévy processes, estimating functions and especially hypothesis testing is a field of current research. Woerner (2001) establishes local asymptotic normality results for estimation purposes in the general case, and Far (2000) studies a restricted model including a compound Poisson process combined with a Brownian motion.

We adopt the following setup. For a fixed time horizon $T > 0$ and each integer $N \geq 1$, we are given a sequence of partitions of $[0, T]$: $0 = t_0^{(N)} \leq t_1^{(N)} \leq \cdots \leq t_{k^{(N)}}^{(N)} \leq T$. The number of partitions $k^{(N)}$ tends to infinity as $N \to \infty$. For each $N \geq 1$, the continuous time process X given by Equation (3.2) sampled at $(t_i^{(N)})_{1 \leq i \leq k^{(N)}}$ is denoted by $X^{(N)}$:

$$X_i^{(N)} = X_{t_i^{(N)}} \,, \quad \text{for } i = 0, \ldots, k^{(N)}. \tag{3.4}$$

Throughout, to avoid some technicalities, we will assume an equispaced (deterministic) partition of the observation period $[0, T]$, so that from now on $k^{(N)} = N$ and $t_i^{(N)} = iT/N$, $i = 0, \ldots, N$. Later (see Section 3.6), we indicate how this can be generalised.

Lemma 3.1 *We can represent the discrete time process $X^{(N)}$ in the form*

$$X_i^{(N)} \;=\; (1 - \alpha_N)\, m \;+\; \alpha_N \, X_{i-1}^{(N)} \;+\; \sigma_N \, \varepsilon_i^{(N)} \,, \quad \textit{for } i = 1, ..., N, \tag{3.5}$$

where $\alpha_N = \alpha_N(\gamma) = e^{-\gamma T/N}$, $\sigma_N = \sigma_N(\gamma) > 0$ and the sequence $\left(\varepsilon_i^{(N)}\right)_{i=1,\ldots,N}$ is a family of independent random variables with zero mean and unit variance, for each fixed $N \geq 1$.

Equation (3.5) represents the continuous time OU process as a kind of discrete time OU or autoregressive process, an analogy which is well known.

Here, we utilize a pseudo-likelihood ratio (LR) test (assuming the residuals are normally distributed in order to set up the likelihood, then dropping this assumption once the estimating equations/estimates are obtained) for the presence/absence of a mean reversion effect. Within the given OU setting, the likelihood estimation is carried out both for the discrete and continuous time models. As in Equation (3.3), from Equation (3.5) we see that this necessitates attention to two unusual features: under the null $\mathbf{H}_0$ the mean reversion parameter m disappears, and furthermore the autoregression becomes unit root (random walk) type, since $\alpha_N = 1$ when $\gamma = 0$. The unit root aspect is well understood now and easily handled by functional limit theorems. We rely on the methods of Davies (1977, 1987) to deal with the fact that the mean reversion level m cannot be identified under the null hypothesis.

3.3 Likelihood Analysis

In this section, we analyze the likelihood function and derive the likelihood ratio (LR) statistic for testing for a mean reversion effect.

Assuming the $\varepsilon_i^{(N)}$ in (3.5) are standard normal, the log-likelihood function of $X_1^{(N)}, \ldots, X_N^{(N)}$ reads (apart from a constant)

$$\mathcal{L}_T^{(N)}(\sigma^2, \gamma, m) \;=\; -\frac{N}{2} \ln\left(\sigma_N^2\right) - \frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{\left[\left(X_i^{(N)} - m\right) - \alpha_N \left(X_{i-1}^{(N)} - m\right)\right]^2}{\sigma_N^2} \right\}, \quad (3.6)$$

where, as in (3.5), $\alpha_N = e^{-\gamma T/N}$, and we define $\sigma_N^2 = \sigma_N^2(\gamma)$ by

$$\sigma_N^2 = \sigma^2 \left(\frac{1 - e^{-2\gamma T/N}}{2\gamma} \right). \quad (3.7)$$

(For $\sigma_N^2(0)$ we take $\sigma^2 T/N$.) At this stage we drop the assumption that the $\varepsilon_i^{(N)}$ are normally distributed, and revert to the former setup, but we still maximise (3.6) to obtain estimators and test hypotheses.

Under the null hypothesis, the disappearance of the nuisance parameter m results in a singular information matrix and standard methods of maximization of the likelihood function fail. We adopt the methodology of Davies (1977, 1987) to deal with this problem.

First, collect the parameters into a 3–vector $\theta = (m, \gamma, \sigma^2)$. When estimating θ under the alternative, we maximize the likelihood on the restricted space

$$\Omega_m = \{m\} \times \mathbb{R} \times \mathbb{R}^+ .$$

As well, we perform an estimation on the restricted null space given by

$$\tau_m = \{m\} \times \{0\} \times \mathbb{R}^+ .$$

The parameter m vanishes from $\mathcal{L}_T^{(N)}$ on τ_m, see Equation (3.6). The restricted deviance $d_{T,m}(N)$, i.e. twice the log likelihood ratio, is

$$d_{T,m}(N) \;=\; -2 \left(\mathcal{L}_T^{(N)} \left(\hat{\theta}_0^{(N)}; m\right) - \mathcal{L}_T^{(N)} \left(\hat{\theta}_m^{(N)}; m\right)\right), \; m \in \mathbb{R}, \quad (3.8)$$

where $\hat{\theta}_0^{(N)}$ and $\hat{\theta}_m^{(N)}$ are the maximizers of $\mathcal{L}_T^{(N)}$ on τ_m and Ω_m respectively.

In the next step, we relax the restriction of the fixed parameter m by taking the sup over all admissible values. This procedure defines the deviance $d_T(N)$ for our specification:

$$d_T(N) = \sup_{m \in \mathbb{R}} d_{T,m}(N). \tag{3.9}$$

The following lemma give us a closed form expression for $d_{T,m}(N)$ as a functional of the observed discrete time process $X^{(N)}$, showing incidentally that it is finite almost surely (a.s.), even though we allow maximisation over all $m \in \mathbb{R}$. We obtain a quite explicit expression for the deviance (see (3.32) and (3.37) below) and (in the next section) for its limit in distribution under the two limiting regimes we consider. To specify these, we need to embed the process $X^{(N)}$ again into continuous time by the adapted piecewise constant càdlàg process

$$X_t^{(N)} = X_{i-1}^{(N)} = X_{t_{i-1}^{(N)}}\,, \quad \text{for } t \in [t_{i-1}^{(N)}, t_i^{(N)}) \text{ and } i = 1, ..., k^{(N)}\,.$$

Lemma 3.2 *Let X be given according to Equation (3.1) and let $X^{(N)}$ be the discrete time sampled version of X for fixed T defined by Equation (3.4), with the partition equispaced, i.e., $k^{(N)} = N$ and $t_i^{(N)} = iT/N$. Then*

$$d_{T,m}(N) = -N \ln\left(1 - \frac{1}{N}\, Z_{T,m}(N)\right), \qquad a.s.\,, \tag{3.10}$$

where $Z_{T,m}(N)$, $m \in \mathbb{R}$, is given by

$$Z_{T,m}(N) = \left(\frac{\left[X^{(N)}, X^{(N)}\right]_T - X_0^2}{T}\right)^{-1} \frac{\left(\int_0^T \left(X_{t-}^{(N)} - m\right) dX_t^{(N)}\right)^2}{\int_0^T \left(X_{t-}^{(N)} - m\right)^2 dt}\,. \tag{3.11}$$

3.4 Asymptotic Results

Our first result gives the asymptotic distribution of $d_T(N)$ as $N \to \infty$ for a refining sequence of partitions. With an increasing number of observations N, we can identify the path of the process X more and more clearly, over a fixed time horizon $T > 0$. This kind of convergence is considered by Boswijk (2001) and Nelson (1990). These authors focus on conditionally heteroscedastic time series models and their continuous time diffusion limits, and Boswijk (2001) allows for mean reversion in addition to conditional heteroscedasticity. General results on continuous time processes as limits of discrete time processes can be found in Kurtz and Protter (1991).

Using Lemma 3.2, we establish an asymptotic result for equispaced partitions that holds for all Lèvy driven OU processes fitting the present framework.

Theorem 3.3 *Let X be given according to Equation (3.1) and let $X^{(N)}$ be the discrete time sampled version of X for fixed T defined by Equation (3.4). Assume the partition is equispaced. Then, under the null hypothesis $\mathbf{H}_0$: $\gamma_0 = 0$, as $N \to \infty$*

$$d_T(N) \xrightarrow{P} \left(\frac{[L,L]_T}{T}\right)^{-1} \left(\frac{\left(\int_0^T L_{t-}\,dL_t - L_T T^{-1}\int_0^T L_t\,dt\right)^2}{\int_0^T L_t^2\,dt - T^{-1}\left(\int_0^T L_t\,dt\right)^2} + \frac{L_T^2}{T}\right), \tag{3.12}$$

If $L = W$ is an SBM we have

$$d_T(N) \xrightarrow{D} \frac{\left(\int_0^1 W_t\,dW_t - W_1\int_0^1 W_t\,dt\right)^2}{\int_0^1 W_t^2\,dt - \left(\int_0^1 W_t\,dt\right)^2} + W_1^2, \quad \text{as } N \to \infty. \tag{3.13}$$

Remark. (1) The convergence in probability in (3.12) is changed to convergence in distribution in (3.13), since we use the self-similarity property of Brownian Motion in the proof. Note that the righthand side of (3.13) does not depend on T.

(2) The result in (3.13) is closely connected to the Dickey-Fuller statistic for testing an autoregressive time series of order one with an included constant term for reduction to random walk, see, e.g., Hamilton (1994), pp. 490. But in addition to the Dickey-Fuller statistic, our asymptotic distribution of the deviance contains the chi-square term W_1^2.

Theorem 3.3 is based on an approximation of the continuous time OU process by appropriately discretized versions. The asymptotic distribution of the deviance involves the driving Lévy process in (3.12). Next we show that, for increasing observation time T, the asymptotic distribution of the deviance does not depend on the exact form of the Lévy process, but can be expressed as the functional of Brownian Motion in (3.13).

Theorem 3.4 *In the situation of Theorem 3.3, let d_T denote the righthand side of (3.12). Then we have $d_T \xrightarrow{D} d$, as $T \to \infty$, where d is a random variable with the same distribution as the random variable on the righthand side of (3.13).*

3.5 From Discrete to Continuous Time

In the previous sections, we discretised the continuous time process and used the approach of Davies (1977, 1987) – in discrete time – for testing under non-standard conditions. In this section, we use an analogous approach in continuous time to show how similar results can be obtained without needing to discretise. We have to restrict ourselves to SBM as the driving Lèvy process, however. In this case, guided by the discrete time analysis, we obtain a completely explicit expression for the distribution of the LR statistic. Throughout the section, we refer to Liptser and Shiryaev (1978, Ch. 17) for maximum likelihood estimation for diffusion-type processes; see also Heyde (1997).

In Section 3.2 we introduced the OU type process given by the SDE in Equation (3.1). Here, we restrict the model to diffusions, hence

$$dX_t = g_\theta(X_t)\,dt + \sigma dW_t^\theta\,, \quad \text{for } 0 \le t \le T, \tag{3.14}$$

where $X_0 \in \mathbb{R}$, $g_\theta : \mathbb{R} \mapsto \mathbb{R}$ is given by $g_\theta(x) = \gamma\,(m - x)$, $\theta = (m, \gamma, \sigma^2)$, and W^θ is an SBM with respect to a measure P_θ on $(\mathcal{S}, \mathcal{F})$. In the following, we outline maximum likelihood estimation procedures for our diffusion-type processes, and work out the likelihood ratio test for the null hypothesis of absence of mean reversion, $\mathbf{H}_0$: $\gamma_0 = 0$.

The collection of different model equations parameterized by $\theta \in \Omega$ is associated with the set of probability measures $\mathcal{P}(\Omega) = \{P_\theta : \theta \in \Omega\}$. Loosely speaking, maximizing the likelihood corresponds to the maximization of the density $dP_\theta(\omega)$ for the observed path $X(\omega)$ by picking the appropriate $\theta \in \Omega$. Fixing a reference ("true") measure $P_{\theta_0} \in \mathcal{P}(\Omega)$, this can be formalized by introducing

$$L_T(\theta) = \frac{dP_\theta}{dP_{\theta_0}}\,, \quad \text{for } \theta \in \Omega. \tag{3.15}$$

The maximum likelihood estimator $\hat{\theta}_T$ will then be given by

$$\hat{\theta}_T = \arg\max_{\theta\in\Omega} L_T(\theta)\,. \tag{3.16}$$

Choosing $\theta_0 = (m_0, 0, \sigma_0^2)$ to specify the reference measure, Equation (3.14) gives $dX_t = \sigma_0\, dW_t^{\theta_0}$. If we define (according to Liptser and Shiryaev (1978, Ch. 17.2))

$$L_T(\theta) = \exp\left(\int_0^T \psi_\theta(X_t)\,dX_t - \frac{\sigma_0^2}{2}\int_0^T \psi_\theta^2(X_t)\,dt\right)\,,$$

with

$$\psi_\theta(x) = \frac{g_\theta(x)}{\sigma_0^2} = \gamma\,\frac{m-x}{\sigma_0^2}\,,$$

then the process $W^\theta = W^{\theta_0} - \sigma_0 \int_0^\cdot \psi_\theta(X_t)\,dt$ is an SBM with respect to the measure P_θ given in (3.15). Accordingly, we recover Equation (3.14), i.e., under P_θ, the process X is of OU type with parameters m and γ. But the parameter σ is forced to be equal to the arbitrarily chosen σ_0, since this cancels out by the definition of ψ_θ. This results from the well-known fact that one may alter the drift of a Brownian Motion but not the scale (volatility) when changing to an equivalent measure. For simplicity, we may assume the true parameter is given by $\sigma_0 = 1$, and hence $\Omega = \mathbb{R} \times \mathbb{R}_0^+$ is the suitably reduced parameter space.[1] Moreover, we set $\theta = (\theta_1, \theta_2) = (m, \gamma)$, and $\theta_0 = (m_0, 0)$, for arbitrary m_0, and introduce the log-likelihood function

$$\mathcal{L}_T(\theta) = \log\left(L_T(\theta)\right) = \int_0^T g_\theta(X_t)\,dX_t - \frac{1}{2}\int_0^T g_\theta^2(X_t)\,dt\,, \tag{3.17}$$

[1] In fact, the scale is not a matter of estimation, since it is a unique path property of the process X. By setting $\hat{\sigma}^2 = [X, X]_t/t$, we have for any arbitrarily small $t > 0$ the P-a.s. identity $\hat{\sigma}^2 = \sigma_0^2$.

where, now, $g_\theta(x) = \gamma\ (m - x)$. The first and negative second derivatives of $\mathcal{L}_T(\theta)$ are

$$\mathcal{S}_T(\theta) = \frac{\partial \mathcal{L}_T}{\partial \theta}(\theta) \quad \text{and} \quad \mathcal{F}_T(\theta) = -\frac{\partial^2 \mathcal{L}_T}{\partial \theta^2}(\theta)\,. \tag{3.18}$$

Estimation and testing can now be performed similarly to the discrete time maximum likelihood. We maximize $\mathcal{L}_T(\theta)$ by solving $\mathcal{S}_T(\theta) = 0$. As for discrete time, the parameter m disappears under the null, when the true parameter satisfies $\gamma_0 = 0$. Correspondingly, the (observed) information matrix becomes singular under $\mathbf{H}_0$:

$$\mathcal{F}_T(\theta_0) = \begin{pmatrix} 0 & -X_T \\ -X_T & \int_0^T (m_0 - X_t)^2 dt \end{pmatrix}.$$

Hence, the entire parameter vector θ cannot be estimated by maximum likelihood when the true process follows Brownian Motion. To test for reduction to this case, we can mimic the approach of Davies (1977, 1987) as given in the discrete time setting.

First, fix m and maximize $\mathcal{L}_T$ on the restricted alternative $\Omega_m = \{m\} \times \mathbb{R}^+$ and on the restricted null hypothesis $\tau_m = \{m\} \times \{0\}$. From (3.17), it is obvious that $\mathcal{L}_T(\theta) = 0$ for all $\theta \in \tau_m$, given an arbitrary $m \in \mathbb{R}$. This holds because we have chosen the reference P_{θ_0} such that X is a Brownian Motion under P_{θ_0}. From $\mathcal{S}_T(\theta) = 0$ we easily obtain, on τ_m,

$$\hat{\gamma}_T(m) = \frac{\int_0^T (m - X_t)\, dX_t}{\int_0^T (m - X_t)^2\, dt}\,, \quad \text{for all } m \in \mathbb{R} \tag{3.19}$$

(c.f. Liptser and Shiryaev (1978), Equation (17.25)). The maximized log-likelihood function on Ω_m reads then

$$\mathcal{L}_T\left(m, \hat{\gamma}_T(m)\right) = \frac{1}{2}\, \frac{\left(\int_0^T (m - X_t)\, dX_t\right)^2}{\int_0^T (m - X_t)^2\, dt}\,, \quad \text{for all } m \in \mathbb{R}. \tag{3.20}$$

Since $\mathcal{L}_T = 0$ on τ_m, the restricted deviance $d_T(m)$ is just two times the maximal log-likelihood function on the restricted alternative:

$$d_T(m) = -2\ \left(\mathcal{L}_T\left(m, 0\right) - \mathcal{L}_T\left(m, \hat{\gamma}_T(m)\right)\right) = \frac{\left(\int_0^T (m - X_t)\, dX_t\right)^2}{\int_0^T (m - X_t)^2\, dt}. \tag{3.21}$$

As in discrete time, we can now define the continuous time deviance d_T as the supremum of the restricted ones:

$$d_T = \sup_{m \in \mathbb{R}} d_T(m)\,. \tag{3.22}$$

We can find the value of this quite explicitly as we do for the discrete time version in Section 2.7 below. Thus we arrive at the following theorem:

Theorem 3.5 *Let X be given according to Equation (3.14). The deviance d_T given by (3.22) is finite a.s. for each T, and, under the null hypothesis $\mathbf{H}_0$: $\gamma_0 = 0$, we have $d_T \stackrel{D}{=} d$, where d is a random variable with the same distribution as the random variable on the righthand side of (3.13).*

Theorem 3.5 shows how to perform an LR test in a diffusion-type setting, when we have to deal with a nuisance parameter under the null hypothesis. In more general situations, we expect that the methodology of Davies (1977) can be applied at least to linear diffusions, and that his approach can be extended to a Lévy process framework. We leave this extension to a future time, but now discuss briefly two extensions in other directions.

3.6 Other Extensions

Various other kinds of extensions of our results are possible. We merely outline here a couple of possibilities.

(1) *Stochastic Volatility.* The OU diffusion X with stochastic volatility V is defined by

$$dX_t = g_\theta(X_t)\,dt + V_t^{\frac{1}{2}} dW_t^\theta = \gamma\,(m-x) + V_t^{\frac{1}{2}} dW_t^\theta\,, \quad \text{for } 0 \le t \le T, \tag{3.23}$$

where $X_0 \in \mathbb{R}$, $g_\theta(x) = \gamma\,(m-x)$, and V is a diffusion-type process, independent of W^θ, given by an SDE

$$dV_t = h_\theta(V_t)\,dt + \sigma_\theta(V_t)\,dZ_t^\theta\,, \quad \text{for } 0 \le t \le T. \tag{3.24}$$

Here h and σ are "well-behaved" functions and Z^θ is an SBM independent of W^θ. The parameter vector θ is of the form $\theta = (\theta^X, \theta^V)$, where $\theta^X = (m, \gamma)$, and θ^V describes the dynamics of V.

As an example, to reproduce a certain continuous time limit of a GARCH(1,1) model as derived by Boswijk (2001) and Nelson (1990), we would define $h_\theta(x) = \lambda\,x + \omega$ and $\sigma_\theta(x) = \xi\,x$, where $\lambda > 0$, $\omega \in \mathbb{R}$, and $\xi > 0$. Then Equation (3.24) reads

$$dV_t = (\lambda V_t + \omega)\;dt + \xi\,V_t\,dZ_t^\theta\,, \quad \text{for } 0 \le t \le T,$$

and we take $\theta^V = (\lambda, \omega, \xi)$.

The solution of the SDE (3.23) is

$$X_t = m + (X_0 - m)\,e^{-\gamma t} + \int_0^t e^{\gamma(s-t)}\,V_s^{\frac{1}{2}}\,dW_s^\theta\,, \quad \text{for } 0 \le t \le T, \tag{3.25}$$

and we can apply continuous time maximum likelihood to the process (X, V).

This results in an expression for the distribution of the deviance that extends Theorem 3.5, and coincides with Theorem 2 of Boswijk (2001):

$$d_T \;\stackrel{D}{=}\; \frac{\left(\int_0^T X_t\,V_t^{-1} dX_t - \left(\int_0^T V_t^{-1} dX_t\right)\,\left(\int_0^T X_t\,V_t^{-1}\,dt\right) / \int_0^T V_t^{-1}\,dt\right)^2}{\int_0^1 X_t^2\,V_t^{-1}\,dt - \left(\int_0^1 X_t\,V_t^{-1}\,dt\right)^2 / \int_0^T V_t^{-1}\,dt} + \frac{\left(\int_0^T V_t^{-1} dX_t\right)^2}{\int_0^T V_t^{-1}\,dt}. \tag{3.26}$$

Further details of this and related analyses will appear elsewhere.

(2) *Random Subintervals.* In (3.4), take $t_i^{(N)}$ to be stopping times with respect to X_t with $\Delta_i^{(N)} = t_i^{(N)} - t_{i-1}^{(N)}$ i.i.d. and positive a.s. for each N, so that $t_i^{(N)}$ is a renewal sequence. Suppose $\mathbb{E}\{\Delta_1^{(N)}\} = T/k^{(N)}$, and $k^{(N)}\mathrm{Var}(\Delta_1^{(N)}) \to 0$ as $n \to \infty$, so we can consider $\{t_i^{(N)}\}_{i=0,\dots,k^{(N)}}$ as a random subinterval of $[0,T]$. Then we can show that the result of Lemma 2.1 remains true, and follow the analysis of Sections 2.3 and 2.7 to obtain a result analogous to that of Theorem 4.1, but so far only under some rather restrictive conditions, such as that m is in a bounded interval. Although this kind of restriction is often made, especially in econometric analyses, we have opted for omitting the quite technical analysis needed for this, preferring to state the results of Sections 2.3 and 2.4 in a clean form which clearly reveals the major issues.

3.7 Simulations, using the VG Process

We conclude with an illustration of the results of the previous sections. We choose the variance gamma (VG) process as driving Lévy process, where we follow Madan and Seneta (1990) and Madan et al. (1998). The aim is to examine how the result of Theorem 3.3 is affected by different choices of Lévy processes L. Additionally, we study the asymptotics in the sense of Theorem 3.4, i.e. for mesh sizes relatively close to zero, we increase the sample time window and analyze the asymptotic behaviour.

The VG process L is a Brownian motion B (with constant drift μ_B and volatility σ_B) evaluated at a random time change given by a gamma process γ:

$$L_t = B_{\gamma_t(\nu)} = \mu_B\, \gamma_t(\nu) + \sigma_B\, W_{\gamma_t(\nu)}\,, \tag{3.27}$$

where W is a standard Brownian motion independent of the process $\gamma(\nu)$, and $\gamma(\nu)$ is a gamma process with unit drift and variance ν. We find for the first two moments of L:

$$\mathbb{E}\{L_t\} = \mu_B\, t\,, \quad \text{and} \quad \mathbb{E}\left\{(L_t - \mu_B\, t)^2\right\} = \left(\mu_B^2\nu + \sigma_B^2\right)\, t\,.$$

To meet the assumptions of Section 3.2 – zero mean and unit variance of L – we have to choose $\mu_B = 0$ and $\sigma_B = 1$. This results in third and fourth moments of the form:

$$\mathbb{E}\left\{L_t^3\right\} = 0\,, \quad \text{and} \quad \mathbb{E}\left\{L_t^4\right\} = 3\,\nu\, t + 3\, t^2\,.$$

The distribution function of L is symmetric, hence exhibits no skewness, but the kurtosis is in general larger than the Brownian kurtosis. The variance $\nu > 0$ of the gamma process determines the kurtosis of the VG process L, and by varying it between 0 and 1, we can study the effect of different types of Lévy processes on the asymptotic distribution of the LR statistics according to Theorem 3.3. At time $T = 1$, this gives us a kurtosis lying between 3 ($\nu = 0$, SBM) and 6 ($\nu = 1$), as a maximum.

Quantiles	50%	75%	90%	95%	97.5%	99%	99.5%	99.9%
$\nu = 1$	2.86	4.65	8.00	11.73	17.17	28.25	40.38	90.68
$\nu = 0.5$	3.07	4.83	7.65	10.51	14.53	21.81	29.00	60.16
$\nu = 0.1$	3.33	5.12	7.45	9.32	11.38	14.49	17.66	26.05
$\nu = 0.05$	3.40	5.21	7.49	9.23	11.01	13.60	15.64	21.43
$\nu = 0.01$	3.42	5.27	7.51	9.12	10.72	12.81	14.51	18.46
$\nu = 0$	3.45	5.31	7.53	9.16	10.73	12.72	14.29	17.63

Table 3.1: Quantiles of the asymptotic distribution of the LR statistics in Eq. (3.12) under the null hypothesis of pure Lévy, where the Lévy process is of VG-type with variance ν.

The results in Table 3.1 are based on 100,000 Monte Carlo runs on equispaced partitions with mesh size $\Delta t = 0.001$. Especially for the high kurtosis case $\nu = 1$, the tail of the distribution becomes very heavy compared to the Brownian case ($\nu = 0$). At the 95% level the quantiles are much larger than the Brownian quantiles, see Rows 1 and 6 in Table 3.1. Nevertheless, for financial data the Brownian quantiles may be reasonable approximations. Madan et al. (1998) investigate the log-return time-series of the S&P 500 index from Jan. 92 to Sep. 94. They find empirical evidence for a symmetric VG type process with $\nu_{S\&P500} = 0.002$, implying a daily kurtosis of 5.19. For this specific data the quantiles/critical values for common testing levels of 90%, 95% and 99% are situated inside the quite narrow intervals given by Rows 5 and 6 of Table 3.1, since $\nu_{S\&P500} = 0.002 \in [0, 0.01]$.

Quantiles	50%	75%	90%	95%	97.5%	99%	99.5%	99.9%
$T = 0.1$	2.86	4.57	7.73	11.32	16.53	26.10	36.57	74.50
$T = 0.5$	3.24	5.01	7.46	9.64	12.10	16.31	20.25	33.70
$T = 1$	3.33	5.12	7.45	9.32	11.38	14.49	17.66	26.05
$T = 5$	3.43	5.28	7.48	9.14	10.95	13.13	14.96	19.19
$T = 10$	3.45	5.31	7.52	9.13	10.82	13.00	14.64	18.72
$T = \infty$	3.45	5.31	7.53	9.16	10.73	12.72	14.29	17.63

Table 3.2: Quantiles of the asymptotics of the LR statistics in Eq. (3.12) under the null hypothesis of pure Lévy, where the Lévy process is of VG-type with $\nu = 0.1$.

Table 3.2 illustrates the asymptotic behaviour of the deviance given in Theorem 3.4 for a VG type process with variance parameter $\nu = 0.01$. Again, the displayed values are based on 100,000 Monte Carlo runs on equispaced partitions with mesh size $\Delta t = 0.001$. For increasing time horizon T, the distribution of the deviance for the VG type process with fixed variance parameter $\nu = 0.1$ converges to the asymptotic distribution given by Equation (3.13). At the 95% significance level, the approximation is already very good for $T \geq 5$.

3.8 Proofs

Proof of Lemma 3.1: By Equation (3.2) and $t_i^{(N)} - t_{i-1}^{(N)} = T/N$, we observe that

$$\begin{aligned}\Delta X_i^{(N)} &:= X_{t_i^{(N)}} - X_{t_{i-1}^{(N)}} \\ &= \left(1 - e^{-\gamma T/N}\right)\left(m - X_{i-1}^{(N)}\right) + \sigma \int_0^{T/N} e^{\gamma(s-T/N)}\, dL_{s+t_{i-1}^{(N)}}\,.\end{aligned}$$

Denoting the last term by $\sigma_N \varepsilon_i^{(N)}$, where σ_N^2 is defined in (3.7), gives the representation in Equation (3.5). Since L is a martingale, $\left(\varepsilon_i^{(N)}\right)_{i=1,\ldots,N}$ are independent and identically distributed with expectation zero. The unit variance of $\varepsilon_i^{(N)}$ follows from

$$\begin{aligned}\mathrm{Var}\left(\int_0^{T/N} e^{\gamma(s-T/N)}\, dL_{s+t_{i-1}^{(N)}}\right) &= \mathbb{E}\left\{\int_0^{T/N} e^{2\,\gamma(s-T/N)}\, d[L,L]_{s+t_{i-1}^{(N)}}\right\} \\ &= \int_0^{T/N} e^{2\,\gamma(s-T/N)}\, ds \\ &= \left(\frac{1 - e^{-2\,\gamma T/N}}{2\,\gamma}\right),\end{aligned}$$

since we have $\mathbb{E}\left\{[L,L]_t\right\} = \mathrm{Var}(L_t) = t$. □

Proof of Lemma 3.2: We have to maximise $\mathcal{L}_T^{(N)}(\sigma^2, \gamma, m)$, as given by (3.6), for variations in σ^2 and γ, for each m. Rather than do this directly we reparameterise according to the 1–1 transformation $(\sigma^2, \gamma) \to (\sigma_N^2, \alpha_N)$ specified by (3.7) and $\alpha_N = e^{-\gamma T/N}$, and maximise $\mathcal{L}_T^{(N)}$ for variations in σ_N^2 and α_N, for each m. We can then find the maximum value of $\mathcal{L}_T^{(N)}$ by substitution of the estimates (which turn out to be uniquely defined) of σ_N^2 and α_N.

Following this program, the maximizers $\hat{\theta}_{T,m}^{(N)}$ and $\hat{\theta}_{T,0}^{(N)}$ of the log-likelihood function in Eq. (3.6) have components given by

$$\begin{aligned}\hat{\alpha}_{T,m}(N) &= \frac{\sum_{i=1}^N \left(X_i^{(N)} - m\right)\left(X_{i-1}^{(N)} - m\right)}{\sum_{i=1}^N \left(X_{i-1}^{(N)} - m\right)^2}\,, \\ \hat{\sigma}_{T,m}^2(N) &= \frac{1}{N}\sum_{i=1}^N \left(\left(X_i^{(N)} - m\right) - \hat{\alpha}_{T,m}^{(N)}\left(X_{i-1}^{(N)} - m\right)\right)^2\,, \quad \text{and} \\ \hat{\sigma}_{T,0}^2(N) &= \frac{1}{N}\sum_{i=1}^N \left(X_i^{(N)} - X_{i-1}^{(N)}\right)^2 = \frac{1}{N}\sum_{i=1}^N \left(\Delta X_i^{(N)}\right)^2.\end{aligned}$$

Substituting these estimators into the log-likelihood function, we obtain for the restricted deviance:

$$\begin{aligned}d_{T,m}(N) &= -2\left(\mathcal{L}_T^{(N)}\left(\hat{\theta}_{T,0}^{(N)}; m\right) - \mathcal{L}_T^{(N)}\left(\hat{\theta}_{T,m}^{(N)}; m\right)\right) \\ &= -N\,\ln\left(\frac{\hat{\sigma}_{T,m}^2(N)}{\hat{\sigma}_{T,0}^2(N)}\right)\end{aligned}$$

$$= \quad -N \ln\left(1 - \frac{1}{N} Z_{T,m}(N)\right) ,$$

where

$$Z_{T,m}(N) = N\left(\frac{\hat{\sigma}^2_{T,0}(N) - \hat{\sigma}^2_{T,m}(N)}{\hat{\sigma}^2_{T,0}(N)}\right) . \tag{3.28}$$

We express $Z_{T,m}(N)$ as a functional of the process $X^{(N)}$ in two steps. First, observe that

$$N\hat{\sigma}^2_{T,0}(N) \;=\; \sum_{i=1}^{N}\left(\Delta X_i^{(N)}\right)^2 \;=\; \left[X^{(N)}, X^{(N)}\right]_T - X_0^2 , \tag{3.29}$$

where we use the bracket notation to denote the quadratic covariation process, see e.g. Protter (1990, p. 58). On the other hand, we have

$$\begin{aligned}
& N\left(\hat{\sigma}^2_{T,0}(N) - \hat{\sigma}^2_{T,m}(N)\right) \\
&= \sum_{i=1}^{N}\left(\left(X_i^{(N)} - X_{i-1}^{(N)}\right)^2 - \left(\left(X_i^{(N)} - m\right) - \hat{\alpha}_{T,m}^{(N)}\left(X_{i-1}^{(N)} - m\right)\right)^2\right) \\
&= \sum_{i=1}^{N}\Big(2\,(\hat{\alpha}_{T,m}^{(N)} - 1)\left(X_i^{(N)} - X_{i-1}^{(N)}\right)\left(X_{i-1}^{(N)} - m\right) \\
&\qquad -(\hat{\alpha}_{T,m}^{(N)} - 1)^2\left(X_{i-1}^{(N)} - m\right)^2\Big) ,
\end{aligned}$$

and by applying

$$\begin{aligned}
\hat{\alpha}_{T,m}(N) - 1 &= \frac{\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)\left(X_i^{(N)} - X_{i-1}^{(N)}\right)}{\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)^2} \\
&= \frac{\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)\Delta X_i^{(N)}}{\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)^2},
\end{aligned}$$

we can conclude

$$\begin{aligned}
N\left(\hat{\sigma}^2_{T,0}(N) - \hat{\sigma}^2_{T,m}(N)\right) &= \frac{\left(\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)\left(X_i^{(N)} - X_{i-1}^{(N)}\right)\right)^2}{\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)^2} \\
&= \frac{\left(\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)\Delta X_i^{(N)}\right)^2}{\sum_{i=1}^{N}\left(X_{i-1}^{(N)} - m\right)^2} \\
&= \frac{\left(\int_0^T\left(X_{t-}^{(N)} - m\right)\, dX_t^{(N)}\right)^2}{(N/T)\int_0^T\left(X_{t-}^{(N)} - m\right)^2\, dt}.
\end{aligned} \tag{3.30}$$

Putting Equation (3.29) and Equation (3.30) together, gives us

$$Z_{T,m}(N) = \left(\frac{\left[X^{(N)}, X^{(N)}\right]_T - X_0^2}{T}\right)^{-1}\frac{\left(\int_0^T\left(X_{t-}^{(N)} - m\right)\, dX_t^{(N)}\right)^2}{\int_0^T\left(X_{t-}^{(N)} - m\right)^2\, dt} , \tag{3.31}$$

and the proof is finished. □

Proof of Theorem 3.3: The proof is divided into three parts. First, we find a closed form expression for the deviance $d_T(N) = \sup_{m\in\mathbb{R}} d_{T,m}(N)$, Part (a). Then we deduce the limiting distribution of $d_T(N)$ for N tending to infinity, Part (b). In Part (c) we specialise to $L = W$, an SBM.
(a) Applying Lemma 3.2 and defining

$$f_N(x) = -N \,\ln\left(1 - \frac{x}{N}\right), \quad \text{for } x \leq N \text{ and } N \in \mathbb{N},$$

we can write $d_{T,m}(N) = f_N\left(Z_{T,m}(N)\right)$. By definition of $Z_{T,m}(N)$ in Equation (3.11) and due to the inequality $0 \leq \hat{\sigma}^2_{T,m}(N) \leq \hat{\sigma}^2_{T,0}(N)$, we observe $0 \leq Z_{T,m}(N) \leq N$. Furthermore, the function f_N is increasing and continuous on $[0, N]$, and hence

$$d_T(N) = \sup_{m\in\mathbb{R}} d_{T,m}(N) = \sup_{m\in\mathbb{R}} f_N\left(Z_{T,m}(N)\right) = f_N\left(\sup_{m\in\mathbb{R}} Z_{T,m}(N)\right). \tag{3.32}$$

With this result, the "sup" is shifted into the function f_N and it is sufficient to analyze $\sup_{m\in\mathbb{R}} Z_{T,m}(N)$. Let

$$a_0 = \left(\frac{1}{T}\int_0^T d\left[X^{(N)}, X^{(N)}\right]_t\right)^{-1}, \quad a_1 = \int_0^T X^{(N)}_{t-}\, dX^{(N)}_t\,, \tag{3.33}$$

$$a_2 = \int_0^T dX^{(N)}_t, \quad a_3 = \frac{1}{\sqrt{T}}\int_0^T X^{(N)}_{t-}\, dt\,, \tag{3.34}$$

$$a_4 = \sqrt{T}, \quad a_5 = \int_0^T (X^{(N)}_{t-})^2\, dt - \frac{1}{T}\left(\int_0^T X^{(N)}_{t-}\, dt\right)^2, \tag{3.35}$$

where the dependence on N and T of the coefficients $a_0, ..., a_5$ is suppressed in the notation. Observe that $Z_{T,m}(N)$ can by written as the quotient of two parabolas:

$$Z_{T,m}(N) = a_0\, \frac{\left(a_1 - a_2\, m\right)^2}{\left(a_3 - a_4\, m\right)^2 + a_5}. \tag{3.36}$$

The coefficients $a_0, .., a_5$ are finite, and $a_1, a_5 > 0$, a.s. Hence, the expression in Eq. (3.36) possesses a unique maximizer $m^\star$ given by

$$m^\star = \frac{a_3}{a_4} + \frac{a_2\, a_5}{a_4\, \left(a_2\, a_3 - a_1\, a_4\right)}.$$

Substituting this into Eq. (3.36), we find in terms of the coefficients $a_0, ..., a_5$

$$Z_T(N) = \sup_{m\in\mathbb{R}} Z_{T,m}(N) = Z_{T,m^\star}(N) = a_0\left(\frac{\left(a_1\, a_4 - a_2\, a_3\right)^2}{a_4^2\, a_5} + \frac{a_2^2}{a_4^2}\right), \tag{3.37}$$

and $d_T(N) = f_N\left(Z_T(N)\right)$. $Z_T(N)$ is a functional of the observable process $X^{(N)}$.
(b) Under the null hypothesis $\gamma = 0$, we want to derive the asymptotic distribution of the

deviance statistics $d_T(N)$ for N tending to infinity. Given $\gamma = 0$, the process X simplifies to $X = X_0 + \sigma L$, see Equation (3.3). A corresponding statement holds for the discretized version, i.e. $X^{(N)} = X_0 + \sigma L^{(N)}$, where $L^{(N)}$ is the process L sampled at $\left(t_i^{(N)}\right)_{i=0,\dots,N}$. Substituting in (3.35), under the null hypothesis $\mathbf{H}_0 : \gamma = 0$, we find $Z_T(N) \xrightarrow{P} Z_T$, as $N \to \infty$, where

$$Z_T = \left(\frac{[L,L]_T}{T}\right)^{-1} \left(\frac{\left(\int_0^T L_{t-}\, dL_t - L_T \frac{1}{T}\int_0^T L_t\, dt\right)^2}{\int_0^T L_t^2\, dt - \frac{1}{T}\left(\int_0^T L_t\, dt\right)^2} + \frac{L_T^2}{T}\right). \tag{3.38}$$

This holds because all components of $Z_T(N)$ converge in probability to their corresponding components of Z_T, and furthermore, $Z_T(N)$ and Z_T respectively can be written as continuous functions of these specific components. The convergence of each component in (3.36) can be briefly verified as follows. Since L is continuous from the right, $L_T^{(N)} \xrightarrow{a.s.} L_T$, and hence $\left(L_T^{(N)}\right)^2 \xrightarrow{a.s.} L_T^2$. The "$dt$"-integral convergences follow from Protter (1990), Theorem 17, Ch. II.—even $a.s.$ With integration by parts, we can decompose $\int L_-^{(N)}\, dL^{(N)}$ into the sum of $(L_T^{(N)})^2$ and the quadratic covariation $[L^{(N)}, L^{(N)}]_T$. By virtue of Theorem 23, Ch. II in Protter (1990), the quadratic covariation process converges uniformly on compacts in probability, i.e. $[L^{(N)}, L^{(N)}] \xrightarrow{ucp} [L, L]$. The result of Eq. (3.38) follows.

Observe that the sequence $(f_N)_N$ converges uniformly on compacts to $f = id_{\mathbb{R}_0^+}$, i.e.

$$\sup_{x\in K} |f_N(x) - f(x)| \quad \to \quad 0\,, \quad \text{for } N \to \infty\,, K \subset \mathbb{R}_0^+ \text{ compact.}$$

With this, and $Z_T < \infty$ a.s., and $Z_T(N) \xrightarrow{P} Z_T$, we can extend the statement of Theorem 17.5 in Jacod and Protter (2000) to get

$$d_T(N) \quad = \quad f_N\left(Z_T(N)\right) \quad \xrightarrow{P} \quad f\left(Z_T\right) \quad = \quad Z_T.$$

□

(c) With SBM as driving Lévy process, all we have to do is apply a linear and deterministic time change combined with an appropriate scaling. Define $\widetilde{W}_t = W_{tT}/\sqrt{T}$, for $t \in [0,1]$. The process $\widetilde{W}$ is an SBM on $[0,1]$. Substituting $\widetilde{W}$ into Equation (3.12) finishes the proof. □

Proof of Theorem 3.4: We have weak convergence in $D[0,1]$ of the process $\widetilde{L}_t^T := L_{tT}/\sqrt{T}$, $0 \le t \le 1$, to an SBM, as $T \to \infty$, as follows. Write, for $t \in [0,1]$,

$$\frac{L_{tT}}{\sqrt{T}} = \sqrt{\frac{\lfloor T \rfloor}{T}} \frac{1}{\sqrt{\lfloor T \rfloor}} \sum_{i=1}^{\lfloor t \lfloor T \rfloor \rfloor} (L_i - L_{i-1}) + \frac{L_{tT} - L_{\lfloor t \lfloor T \rfloor \rfloor}}{\sqrt{T}}.$$

Since the increments $L_i - L_{i-1}$ are i.i.d with expectation 0 and variance 1 (recall that $\mathbb{E}L_t^2 = t$), the first term on the righthand side tends weakly to W_t by the functional central limit theorem for random walks. Also,

$$0 \le tT - \lfloor t \lfloor T \rfloor \rfloor \le t + 1 \le 2.$$

Thus the remainder term can be bounded as follows:

$$\frac{\sup_{0\le t\le 1}|L_{tT}-L_{\lfloor t\lfloor T\rfloor\rfloor}|}{\sqrt{T}} \stackrel{D}{=} \frac{\sup_{0\le t\le 1}|L_{tT-\lfloor t\lfloor T\rfloor\rfloor}|}{\sqrt{T}} \le \frac{\sup_{0\le s\le 2}|L_s|}{\sqrt{T}}.$$

This tends to 0 in probability as $T\to\infty$ because $\sup_{0\le s\le 2}|L_s|$ is stochastically bounded (by Doob's inequality, e.g., Bertoin 1996, p.4). Hence we have the required weak convergence.

Substituting $\widetilde{L}^T$ into Equation (3.12) of Theorem 3.3 the statement of the theorem follows, since the weak convergence of the processes $\widetilde{L}^T$ to W implies the joint convergence in distribution of the rv's

$$\left((L_1^T)^2, \int_0^1 L_{t-}^T dL_t^T, \int_0^1 L_t^T dt, \int_0^1 (L_t^T)^2 dt\right) \stackrel{D}{\longrightarrow} \left(W_1^2, \int_0^1 W_t dW_t, \int_0^1 W_t dt, \int_0^1 W_t^2 dt\right);$$

for details see, e.g., Theorem 2.2 or Theorem 2.7 in Kurtz and Protter (1991). □

Proof of Theorem 3.5: Equation (3.21) is of the same form as Equation (3.31), and hence as Equation (3.36). Maximising over m in Equation (3.21) gives the righthand side of Equation (3.37), but with $a_0,\dots,a_5$ defined in terms of W_t rather than $X_t^{(N)}$. This together with $W_t \stackrel{D}{=} W_{tT}/\sqrt{T}$ gives the result. □

Chapter 4

On the Valuation of Employee Share Options

4.1 Introduction

In recent years there has been a dramatic increase in the use of employee stock options (ESOs) in compensation packages offered by firms. This increase has received wide exposure in the media and generated substantial critical controversy over the fair value of the ESO component of these compensation contracts. Professional bodies, such as accounting standards boards, have been investigating ways in which ESOs should be valued and reported. Certain distinctive features of typical ESOs mean that conventional option pricing methodologies are inapplicable.

Two modelling approaches have previously been used to determine the value of an ESO contract. Structural models directly model the employee's utility-maximizing behavior. The assumption is that employees act so as to maximize their expected utility, subject to some hedging restrictions (see Lambert et al. 1991, Huddart 1994, Kulatilaka and Marcus 1994, and Hall and Murphy 2000). This method requires, among other things, knowledge of unobservable factors such as the executive's non-firm-related wealth, his or her holdings of shares in the firm, the executive's risk aversion, diversification desires, liquidity needs and potential gains from voluntary separation. Consequently these models can be criticized on the grounds that estimating their parameters is impractical, and that these estimates may be difficult for an individual to verify independently of the employee to whom the ESO is issued. By contrast, reduced form models assume that the early exercise event/employee's departure can be modelled by some random process, the parameters for which can be estimated using observed variables, see for example Cuny and Jorion (1995) and Carr and Linetsky (2000), both of which illustrate reduced form, intensity-based approaches to ESO valuation.

In the following, we develop a reduced form model for ESO valuation in continuous time, extending Cuny and Jorion (1995), Carr and Linetsky (2000), and Maller et al. (2002). The fundamental components of the setup is the financial market carrying the relevant tradable assets and two random times announcing departure/exercise and takeover. The financial market model consists of the company's stock price process and other price processes that are needed for performance linked options, e.g., an equity index (see Johnson and Tian, 2000), and all price processes are modelled as diffusion processes.

The two random times of departure and takeover are both associated with the first jumps of two different Cox processes that are both completely described by their random intensities, see Lando (1998). We point out that the application of the different Cox processes does not imply that the two random times are independent: The dependency of the random times is specified via the intensities that are both functions of the effective state of the market prices. For example, in a bull market the takeover intensity tends to increase, and so does the departure/exercise intensity, because the increasing departure/exercise intensity represents the employee's desire for liquidity or diversification, see Carr and Linetsky (1998). Thus, in a bull market employee departure and takeover are positively correlated. Moreover, the structure of the postulated financial market model needs to be understood. The reduced form model is frequently applied, though its probabilistic structure is rarely studied. Besides El Karoui and Martellini (2001), and Blanchet-Scalliet and Jeanblanc (2002), we analyze this structure and clarify the assumptions underlying the model, and discuss their implications. By the nature of the construction, the market model is incomplete, since departure time and takeover time are stochastic sources that contribute to the randomness of the market but are not representable by traded assets. This market incompleteness results in a set of pricing systems, i.e. equivalent martingale measures, rather than a single price for a given contingent claim, hence there are a range of fair values spanned by the convex set of equivalent martingale measures. We provide an explicit characterization of the set of equivalent martingale measures and moreover, we compute prominent martingale measures like, e.g., the variance optimal martingale measure and the minimal martingale measure. Particular ESO specifications are studied emphasizing different aspects of the proposed framework. In this context, we also provide strict no-arbitrage bounds for ESO prices by applying optimal stopping. Furthermore, possible limitations of the proposed model are explored by examining departures from the crucial assumptions of no-arbitrage, e.g., by considering the effects of the employee having inside information, see Kusuoka (1999), and Elliot et al. (2000) and, Blanchet-Scalliet and Jeanblanc (2002) for related discussions.

4.2 A Reduced Form Model

The model is set in a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, P)$ where P is some subjective probability measure and $T > 0$ is a fixed time horizon. We assume the filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ satisfies the usual conditions (see Protter, 1990, p. 3), and the initial sigma field $\mathcal{F}_0$ is trivial, and furthermore $\mathcal{F} = \mathcal{F}_T$.

The financial market model is defined as follows:

Definition 4.1 *The financial market is given by a price process $S = (S^0, S^1, ..., S^d)$ that is driven by a d-dimensional process $W = (W^1, ..., W^d)$.*

(a) The market information is accumulated in the filtration $(\mathcal{G}_t)_{0\le t\le T}$ that is a sub-filtration of $(\mathcal{F}_t)_{0\le t\le T}$, i.e. $\mathcal{G}_t \subset \mathcal{F}_t$ for $0 \le t \le T$.

(b) The process S is càdlàg and $(\mathcal{G}_t)_{0\le t\le T}$-adapted, and W is a Standard Brownian Motion w.r.t. $(\mathcal{G}_t)_{0\le t\le T}$.

(c) The bank account is described by S^0 and given by $dS^0_t = r\, S^0_t\, dt$ for some instantaneous rate $r > 0$, and initial value $S^0_0 = 1$.

(d) The stock price processes $S^1, ..., S^d$ are defined by

$$dS^k_t = S^k_t\, \mu^k\, dt + S^k_t \sigma^k\, dW_t\,, \quad \text{for } 0 \le t \le T \text{ and } k = 1, ...d,$$

where $S^1_0, ..., S^d_0 > 0$, $\mu = (\mu^1, ..., \mu^d)^\top \in \mathbb{R}^d$, and $\Sigma = (\sigma^1, ..., \sigma^d) \in \mathbb{R}^{d\times d}$.

The price process S^1 describes the evolution of the stock price of the firm that is granting the ESO. The other assets given by their price processes $S^2, ..., S^d$ are possibly needed for performance linked ESOs. As usual, the market is completed by the riskless asset S^0.

Part (a) and (b) of Definition 4.1 structure the information sets/filtrations. The financial market is given by the price process S on the filtration $(\mathcal{G}_t)_{0\le t\le T}$, where there might be some additional information in the larger filtration $(\mathcal{F}_t)_{0\le t\le T}$ that is not observable on the financial market. Part (c) and (d) define the well-known Black&Scholes model with constant coefficients on the filtration $(\mathcal{G}_t)_{0\le t\le T}$. Thus, the stock price processes form a d-dimensional Geometric Brownian Motion (GBM) in this setting.

Besides the financial market, the proposed reduced form model is also characterized by the random times T^V, T^D and T^{TO}, where T^V denotes the time of vesting for the specific ESO, T^D the time of early departure of the executive, and T^{TO} the time a take over occurs.

Definition 4.2 *The random times T^V, T^D and T^{TO} are associated with their indicator processes:*

$$N^V_t = \mathbf{1}_{\{T^V \le t\}}\ , N^D_t = \mathbf{1}_{\{T^D \le t\}}\ , \text{and } N^{TO}_t = \mathbf{1}_{\{T^{TO} \le t\}} \qquad \text{for } 0 \le t \le T\,.$$

In Definitions 4.1 and 4.2 the reduced form model is outlined, but a number of properties are not specified. For example relations between the filtrations $(\mathcal{G}_t)_{0\le t\le T}$ and $(\mathcal{F}_t)_{0\le t\le T}$ needs to be defined, and as well we have to formulate the nature of the indicator processes N^V, N^D, and N^{TO}.

Assumption 4.3 *The filtrations $(\mathcal{G}_t)_{0\leq t\leq T}$, and $(\mathcal{F}_t)_{0\leq t\leq T}$ are the augmented natural filtrations of the process S, and the processes S, N^V, N^D, N^{TO} respectively.*

(a) The matrix Σ is of full rank d, and the process W is also a Standard Brownian Motion with respect to the filtration $(\mathcal{F}_t)_{0\leq t\leq T}$.

(b) The random time T^V is a stopping time w.r.t. the filtration $(\mathcal{G}_t)_{0\leq t\leq T}$.

(c) The indicator processes N^D and N^{TO} of the random times T^D and T^{TO} admit intensities λ^D and λ^{TO}, i.e.

$$M^D = N^D - \int_0^{\cdot\wedge T^D} \lambda_u^D\, du \quad \textit{and} \quad M^{TO} = N^{TO} - \int_0^{\cdot\wedge T^{TO}} \lambda_u^{TO}\, du$$

are $(\mathcal{F}_t)_{0\leq t\leq T}$-martingales, where $s\wedge t = \inf\{s,t\}$, and furthermore M^D and M^{TO} are strongly orthogonal. Moreover, λ^D and λ^{TO} are $(\mathcal{G}_t)_{0\leq t\leq T}$-predictable and

$$\int_0^T \lambda_t^D\, dt < \infty\,, \;\textit{and}\; \int_0^T \lambda_t^{TO}\, dt < \infty\,, \quad P-a.s.$$

Remark. (1) Part (a) of Assumption 4.3 implies that the financial market $(S, (\mathcal{G}_t)_{0\leq t\leq T})$ is complete, because Σ is of full rank. The property that W remains a martingale after enlargement of the filtration implies the so-called (H) hypothesis: Every $(\mathcal{G}_t)_{0\leq t\leq T}$-square integrable martingale is a $(\mathcal{F}_t)_{0\leq t\leq T}$-square integrable martingale. This result follows directly by applying the predictable representation theorem for Brownian motion, see Protter (1990), Ch. IV. The (H) hypothesis can be found in papers of Jeulin and Yor where the enlargement of filtrations is studied, see Jeulin and Yor (1979). This topic has been revisited for credit risk modelling purposes, see Elliot et al. (2000). The (H) hypothesis is a sufficient no-arbitrage condition for the market $(S, (\mathcal{F}_t)_{0\leq t\leq T})$. Think of the random times T^D and T^{TO} carrying future information that is actually not available on the financial market. The knowledge of this certain information possibly creates arbitrage opportunities. The (H) hypothesis rules out such situations. On the other hand, it should be noted that the (H) hypothesis does not allow the departure of the employee and takeover to affect the return or volatility structure of the financial market, see El Karoui and Martellini (2001).
(2) The vesting time T^V is a stopping time w.r.t. the Brownian filtration $(\mathcal{G}_t)_{0\leq t\leq T}$, see Part (b). The fact that every local $(\mathcal{G}_t)_{0\leq t\leq T}$-martingale is continuous, see Protter (1990), Ch. IV, Corollary 1 of Theorem 42, implies that the indicator process $N^V = \mathbf{1}_{\{T^V\leq\cdot\}}$ admits a Doob-Meyer decomposition allowing for no martingale component. Thus, the compensator of N^V is the process itself which is a noticeable difference to the Doob-Meyer decompositions of the indicator processes of T^D and T^{TO}.
(3) Both martingales, M^D and M^{TO} are strongly orthogonal to the Standard Brownian Motion W driving the financial market. This result is obvious since W is an almost sure continuous martingale and M^D and M^{TO} are quadratic pure jump (for definition see Protter, 1990, Ch. II.), but useful for establishing a martingale representation theorem, see Kusuoka (1999). Moreover, the model $(S, (\mathcal{F}_t)_{0\leq t\leq T})$ is incomplete, since M^D and M^{TO}

are random sources of risk, but not present in tradable assets.
(4) Part (c) of Assumption 4.3 in combination with the (H) hypothesis leads naturally to the Cox process nature of T^D and T^{TO}, see Blanchet-Scalliet and Jeanblanc (2002) for reference. An appealing consequence of the Cox property is

$$P\left(T^D > t \,\middle|\, \mathcal{G}_T\right) = e^{-\int_0^t \lambda_u^D \, du}, \quad \text{for } 0 \le t \le T, \tag{4.1}$$

and unconditional

$$P\left(T^D > t\right) = \mathbb{E}_P\left\{e^{-\int_0^t \lambda_u^D \, du}\right\}, \quad \text{for } 0 \le t \le T. \tag{4.2}$$

Especially, Equation (4.1) highlights the Cox nature: Conditioned on the entire information of the financial market $\mathcal{G}_T$, we can think of T^D as the first jump of an (inhomogeneous) Poisson process with intensity λ^D, or equivalently, conditioned on $\mathcal{G}_T$, the stopping time T^D is exponentially distributed with stochastic intensity λ^D. The same arguments of course hold for T^{TO}, and furthermore $P(T^D = T^{TO}) = 0$, since both indicator processes are strongly orthogonal, see, e.g., example in the Appendix. However, this implies not that the stopping times are independent. T^D and T^{TO} are determined by their intensities, and these intensities both are functions of the state of the market prices. Thus, the dependency of the stopping times is given in terms of their intensities, e.g., in a specific situation/state of market prices the employee departure is likely but takeover appears to be unlikely. This situation is modelled by assigning the intensity of T^D a very high level, and at the same time the intensity of T^{TO} is set close to zero.
(5) The intensity of N^D and N^{TO} in the terminology of point processes and martingale theory are of course $\lambda^D \mathbf{1}_{\{T^D > \cdot\}}$ and $\lambda^{TO} \mathbf{1}_{\{T^{TO} > \cdot\}}$, see, e.g., Brémaud (1981).

Example. In the proposed reduced form model an ESO is a contingent claim within an incomplete market model induced by S, N^D and N^{TO}. For illustration purposes, we restrict the model to $d = 1$, define $\lambda_t^D = h(S_t, t)$ for some measurable function h, and ignore the possibility of a takeover. Furthermore, we set $\mu = r$ and hence S^1/S^0 is a martingale, and $Q = P$ is called risk-neutral probability measure.

By these temporary restrictions we are in the framework of Carr and Linetsky (2000). They describe an ESO for a deterministic vesting date $T^V \in [0, T]$ by an European call option with pre-specified strike price $K > 0$ and maturity T, where they also allow for early departure announced by T^D. The value at $t \in [0, T]$ of an unexercised ESO is given by the risk-neutral expectation:

$$C(S_t, t; K, T) = \mathbb{E}_Q\left\{\mathbf{1}_{\{T^D \ge t \vee T^V\}} e^{-r(T \wedge T^D - t)} \left(S_{T \wedge T^D} - K\right)^+ \mid \mathcal{F}_t\right\}, \text{ for } 0 \le t \le T. \tag{4.3}$$

If the employee's departure takes place before vesting in T^V the option becomes worthless, otherwise the ESO is settled at the time of early departure with the intrinsic value.

For the departure intensity $\lambda_t^D = h(S_t, t)$ Carr and Linetsky (2000) consider explicitly two specifications:

$$h(S_t, t) = \lambda_f + \lambda_e \mathbf{1}_{\{S_t > K\}}, \quad \text{and} \quad h(S_t, t) = \lambda_f + \lambda_e \left(\ln S_t - \ln K\right)^+. \tag{4.4}$$

The first formulation leads to the so-called occupation time specification, whereas the second is denominated Brownian area specification. For both specifications they derive rather explicit expressions for the ESO value. Finally, we remark that the model setup by Carr and Linetsky (2000) is exactly the continuous time version of the discrete time model of Cuny and Jorion (1995).

4.3 The Probabilistic Structure of the Model

The model defined in the previous section is discussed for its characteristics and properties. A fundamental role plays the *predictable representation property.* According to Theorem 2.3 in Kusuoka (1999), this property can be established for the proposed reduced form model, see Appendix for details. The predictable representation theorem, Theorem 4.15, is useful for describing the set of equivalent measures.

Theorem 4.4 (Björk et al., 1997) *Let $\psi = (\psi^1, ..., \psi^d)$ and $\phi = (\phi^D, \phi^{TO})$ be strictly positive predictable processes satisfying P-almost surely*

$$\int_0^T \|\psi_t\|^2 \, dt < \infty \,, \quad \int_0^T \phi_t^D \, \lambda_t^D \, dt < \infty \,, \quad \int_0^T \phi_t^{TO} \, \lambda_t^{TO} \, dt < \infty \,. \tag{4.5}$$

Define the density process L by

$$dL_t = L_{t-} \left(\psi_t \, dW_t + (\phi_t^D - 1) \, dM_t^D + (\phi_t^{TO} - 1) \, dM_t^{TO} \right), \quad \text{for } 0 \le t \le T, \text{ and } L_0 = 1. \tag{4.6}$$

Provided $\mathbb{E}_P\{L_T\} = 1$, the measure Q defined by $dQ = L_T \, dP$ is equivalent to P, and

$$\widetilde{W} = W - \int_0^{\cdot} \psi_t \, dt, \ \widetilde{M}^D = N^D - \int_0^{\cdot \wedge T^D} \phi_t^D \, \lambda_t^D \, dt, \ \widetilde{M}^{TO} = N^{TO} - \int_0^{\cdot \wedge T^{TO}} \phi_t^{TO} \, \lambda_t^{TO} \, dt, \tag{4.7}$$

are Q-martingales, moreover $\widetilde{W}$ is a Q-SBM.
Furthermore, on $(\mathcal{F}_t)_{0 \le t \le T}$ every probability measure $Q \sim P$ can be presented in the way given above, where $(\mathcal{F}_t)_{0 \le t \le T}$ is the augmented natural filtration of W, N^D, and N^{TO}.

Remark. (6) An arbitrary change of measure does usually not preserve the Cox process property (see Remark 4). Proposition 2 in Blanchet-Scalliet and Jeanblanc (2002) requires ψ to be $(\mathcal{G}_t)_{0 \le t \le T}$-adapted to preserve the Cox property/(H) hypothesis. This condition turns out to be sufficient in their case, because their setup incorporates only one stopping time τ, i.e. the time of default: On the set $\{\tau > t\}$, any $\mathcal{F}_t$-measurable rv is equal to some $\mathcal{G}_t$-measurable rv, and hence no further restriction on ϕ is necessary. Kusuoka (1999) gives an example for a setup with two stopping times where the Cox property is not preserved. A sufficient condition for preserving the Cox property under a change of measure is to require ψ and ϕ to be $(\mathcal{G}_t)_{0 \le t \le T}$-adapted.

The Girsanov-type statement in Theorem 4.4 is well known, and frequently applied to term structure and credit risk modelling. We utilize this statement in order to characterize the set of the equivalent martingale measures of the proposed reduced form model.

Theorem 4.5 *Let the reduced form model be given by Definition 4.1, and Definition 4.2, and Assumption 4.3, and denote $\mathcal{Q}$ the set of all equivalent martingale measures. Then every $Q \in \mathcal{Q}$ is given in terms of (ψ, ϕ) by $dQ = L_T^{\psi,\phi}\, dP$, where $L^{\psi,\phi}$ is defined in Equation (4.6), and*

$$\psi_t^\top = -\Sigma^{-1}(\mu - r\,\mathbf{e})\,, \quad \text{for } 0 \le t \le T, \tag{4.8}$$

where $\mathbf{e} = (1, ..., 1)^\top$, and ϕ is arbitrary supposed the conditions of Theorem 4.4 hold.

Following the above theorem, any EMM Q is solely described by ϕ, and therefore, we write $Q^\phi = Q$ for all $Q \in \mathcal{Q}$. Furthermore, the set of EMMs is not a singleton, and hence the market model is not complete. For incomplete market models the valuation of contingent claims, or in our case ESOs, results not into a single price, but in an interval representing the range of fair prices. The following definition describes two subsets of $\mathcal{Q}$ that enable explicit calculations for later purposes.

Definition 4.6 *Following Theorem 4.5, $Q^\phi \in \mathcal{Q}$ is given by $\phi = (\phi^D, \phi^{TO})$.*

(a) $\mathcal{Q}^{Cox} = \{Q^\phi \in \mathcal{Q} \,:\, \phi \text{ is } (\mathcal{G}_t)_{0 \le t \le T}\text{-predictable}\}$.

(b) $\mathcal{Q}^{exp} = \{Q^\phi \in \mathcal{Q} \,:\, \phi^D\,\lambda^D \text{ and } \phi^{TO}\,\lambda^{TO} \text{ are constant}\}$.

The set $\mathcal{Q}^{Cox}$ contains those EMMs that are preserving the Cox property, see Remark (6), and $\mathcal{Q}^{exp}$ collects those elements of $\mathcal{Q}$ such that the intensity of N^D and N^{TO} are constant, and hence T^D and T^{TO} are exponentially distributed stopping times. Clearly, we have the relation

$$\mathcal{Q}^{exp} \subset \mathcal{Q}^{Cox} \subset \mathcal{Q}\,.$$

Next, we recall the definition of two prominent EMMs: the variance optimal martingale measure and the minimal martingale measure, cf. Schweizer (1995). For the latter measure, we apply the characterization in terms of the relative entropy established for continuous price processes by Schweizer (1995), Theorem 5.

Definition 4.7 *Recall that $\mathcal{Q}$ is the set of equivalent martingale measures.*

(a) The measure $Q^{opt} \in \mathcal{Q}$ that minimizes $L^2(P)$ distance to P

$$D(Q, P) = \left\| \frac{dQ}{dP} - 1 \right\|_{L^2(P)} = \mathrm{Var}_P \left(\frac{dQ}{dP} \right)^{1/2} ,$$

is called variance optimal martingale measure.

(b) The measure $Q^{min} \in \mathcal{Q}$ that minimizes the relative entropy

$$H(Q|P) = \int \log \left(\frac{dQ}{dP} \right) dQ\,,$$

is called minimal martingale measure.

The incompleteness of the reduced form model is introduced by the random times announcing employee's early departure T^D and takeover T^{TO}. Both random times are stochastic sources within the model, but are not present in the tradable assets. A very special structure of incomplete market is generated, since the corresponding martingales M^D and M^{TO} are strongly orthogonal to the martingales driving the price process S of the tradable assets. This results into a particular form of the minimal and the variance optimal martingale measure.

Theorem 4.8 *Let the reduced form model be given by Definition 4.1, and Definition 4.2, and Assumption 4.3. Then the variance optimal martingale measure Q^{opt} and the minimal martingale measure Q^{min} coincide, i.e. $Q^{opt} = Q^{min} = Q^\star$. And the measure $Q^\star$ is defined in terms of $(\psi^\star, \phi^\star)$ in Equation (4.6) by $dQ^\star = L^{\psi^\star,\phi^\star}\, dP$, where*

$$\psi_t^\star = -\Sigma^{-1}(\mu - r\,\mathbf{e})\,, \quad \textit{and} \quad \phi_t^\star = (1,1) \quad \textit{for } 0 \le t \le T. \tag{4.9}$$

Moreover, for every $Q \in \mathcal{Q}$ we have an additive decomposition of the relative entropy

$$H(Q|P) = H(Q|Q^\star) + H(Q^\star|P)\,, \tag{4.10}$$

provided $H(Q|P)$ exists.

Remark. (7) For proving Theorem 4.8, we do not need the strict definition of the price process, i.e. deterministic components for the drift and the covariance structure, see Definition 4.1. We can relax this restriction, and the proof still applies as long as the price process S is adapted to natural Brownian filtration $(\mathcal{G}_t)_{0\le t\le T}$.
(8) By similar argumentation as in the proof of Theorem 4.8, we can show that the martingale measure induced by the numeraire portfolio or the Esscher transform also coincides with $Q^\star$. Both EMMs are based on the structure of the price process S, and the market $(S, (\mathcal{G}_t)_{0\le t\le T})$ is complete, so all discussed martingale measures coincide with the unique EMM on $(\mathcal{G}_t)_{0\le t\le T}$ and are embedded in $(\mathcal{F}_t)_{0\le t\le T}$ by virtue of the (H) hypothesis.

4.4 The Valuation of ESOs

Various specifications of ESOs are discussed in this section. Within the proposed reduced form setting, a general description of an ESO is hardly possible, since an ESO is any arbitrary contingent claim in the given framework. Thus, we study specific ESOs that are a collection of distinct examples emphasizing different aspects rather than a complete characterization. For the following discussion, we choose an arbitrary equivalent martingale measure $Q \in \mathcal{Q}$ and furthermore, we give the range of possible fair values. But we point out that the minimal martingale/variance optimal measure $Q^\star$ is a proper choice for the valuation of ESOs.

4.4.1 Takeover Provisions

To illustrate the effect of takeover provisions, we first restrict to the case when $\lambda^D = 0$, i.e. employee's departure does not take place. Furthermore, the vesting date T^V is non-random, hence a priori given, and $T^V \in [0, T[$. By these restrictions, we have one random time $\tau = T^{TO}$. We consider a specific company that is the potential aspirant for a takeover. This company is represented by its stock price process S^2, and we assume that no other price process is necessary for the model specification, therefore $d = 2$ and $S = (S^0, S^1, S^2)$. We remark that the presence of exactly one stopping time τ in the setup implies $\mathcal{Q} = \mathcal{Q}^{Cox}$, i.e. the Cox property holds for all equivalent martingale measures, see Remark (6).

In the above specified setting, we have an pre-vesting takeover, i.e. $\tau < T^V$, and the regular takeover given by $T^V \leq \tau \leq T$. The main difference to the employee departure is that pre-vesting takeover assigns the ESO usually a certain value, whereas early departure forfeits the ESO. Also note that in general a takeover increases the value of an ESO. A possible ESO specification is the contingent claim X with maturity $T \wedge \tau$.

$$X \;=\; \mathbf{1}_{\{\tau<T^V\}}c \,+\, \mathbf{1}_{\{T^V\leq\tau<T\}}\max\{S^1_\tau, S^2_\tau\} \,+\, \mathbf{1}_{\{T\leq\tau\}}(S^1_T - K)^+\,, \tag{4.11}$$

where $c > 0$ is a cash compensation paid in the case of pre-vesting takeover. When takeover takes place after vesting of the ESO, the employee is rewarded with the maximum value of the share price of his employing firm and the company taking over. If no takeover occurs, we have the typical ESO specification, i.e. the employee gets European style call option pay-off with pre-specified strike price K.

To formalize this example, let the Q-intensity of takeover in τ be defined in a time-homogeneous way by $\lambda_t = h(S^1_t, S^2_t)$, where h is a function in $C^2(\mathbb{R}^2, \mathbb{R}^+)$. For the fixed measure $Q \in \mathcal{Q}$, let $C(t, S^1_t, S^2_t; c, K, T)$ denote the price of the contingent claim given in Equation (4.11), and in terms of the expectation we find

$$\begin{aligned} C(t, S^1_t, S^2_t; c, K, T) \;&=\; c\,\mathbb{E}_Q\{e^{-r\,(\tau-t)}\mathbf{1}_{\{\tau<T^V\}}\,|\mathcal{F}_t\} \\ &\quad + \mathbb{E}_Q\{e^{-r\,(\tau-t)}\mathbf{1}_{\{T^V\leq\tau<T\}}\max\{S^1_\tau, S^2_\tau\}\,|\mathcal{F}_t\} \\ &\quad + e^{-r\,(T-t)}\mathbb{E}_Q\{\mathbf{1}_{\{T\leq\tau\}}(S^1_T - K)^+\,|\mathcal{F}_t\}\,, \quad \text{for } 0 \leq t \leq T. \end{aligned}$$

The first line is a similar formula as it is known for default options, see, e.g., Lando (1998), the second line is related to an exchange option with random maturity subject to the condition $\{T^V \leq \tau < T\}$, and the last line is a common European call option subject to the condition $\{\tau \geq T\}$.

The above expression can be simplified by applying the Cox property of τ, i.e. by conditioning on $\mathcal{G}_T$. With $(x \vee y) = \max\{x, y\}$, we find for $t = 0$

$$c\,\mathbb{E}_Q\left\{e^{-r\,\tau}\mathbf{1}_{\{\tau<T^V\}}\right\} \;=\; c\int_0^{T^V}\mathbb{E}_Q\left\{h(S_t)e^{-\int_0^t(r+h(S_u))du}\right\}dt\,, \tag{4.12}$$

$$\mathbb{E}_Q\left\{e^{-r\,\tau}\mathbf{1}_{\{T^V\leq\tau<T\}}(S^1_\tau \vee S^2_\tau)\right\} \;=\; \int_{T^V}^{T}\mathbb{E}_Q\left\{h(S_t)e^{-\int_0^t(r+h(S_u))du}(S^1_t \vee S^2_t)\right\}dt \tag{4.13}$$

$$e^{-r\,T}\mathbb{E}_Q\left\{\mathbf{1}_{\{T\leq T^{TO}\}}(S^1_T - K)^+\right\} \;=\; \mathbb{E}_Q\left\{e^{-\int_0^T(r+h(S_u))du}(S^1_T - K)^+\right\}\,. \tag{4.14}$$

The expectation on the right of Equations (4.12–4.14) can be expressed in terms of solutions of partial differential equations by application of the Feynman-Kac framework, see Karatzas and Shreve (1988). This is possible since the intensity h is defined as a function of the diffusion process S. Especially, Equation (4.14) highlights the analogy of ESO valuation and credit derivative pricing, since this expectation can be interpreted as European call option with stochastic interest rate $r + h$.

For the takeover effect on the ESO structure given by Equation (4.11), the no-arbitrage bounds of all possible fair values can be derived. Observe that τ can be controlled by its intensity λ that can be specified arbitrarily, see Theorem 4.5. If we have a time when takeover would imply maximal/minimal reward for the employee then we can increase the intensity arbitrarily large, and τ occurs. Thus, the non-arbitrage bounds are connected to the optimal stopping problem/American options, see Karatzas and Shreve (1998):

The optimal stopping problem (Y, P) is to find a stopping time $\tau^\star = \tau^\star(\omega, y)$ for the continuous process Y that is bounded from below such that

$$\mathbb{E}^y\{Y_{\tau^\star}\} = \sup_{\tau \in \mathcal{S}_{0,T}} \mathbb{E}^y\{Y_\tau\},$$

where $\mathbb{E}^y$ denotes the expectation under P with initial condition $Y_0 = y$, $\mathcal{S}_{0,T}$ is the set of stopping times taking values in $[0, T]$, and the underlying filtration $(\mathcal{F}_t)_{0 \le t \le T}$ is given by the natural filtration of Y.

Proposition 4.9 *In the present setting, let an ESO be given by the process* $X_t = F(t, S_t)$, *where the pay-off in* τ *is* X_τ *if* $\tau \le T$, *respectively* X_T *if* $\tau > T$, *and* F *is measurable function such that* X *is continuous and bounded from below. Then*

$$\mathbb{E}_Q\{e^{-r\,(\tau\wedge T)}\, X_{\tau\wedge T}\} \;\le\; \mathbb{E}_{Q^\star}\{e^{-r\,\tau^\star}\, X_{\tau^\star}\}\,, \quad \text{for all } Q \in \mathcal{Q}\,, \tag{4.15}$$

where $\tau^\star$ *is the solution of the optimal stopping problem* $(Y, Q^\star)$, *and* Y *is given by* $Y = X/S^0$. *Furthermore, the above equation gives a strict upper boundary for the ESO price.*

Remark. (9) In Proposition 4.9, if we additionally suppose X is bounded from above, then we can establish a strict lower bound of the ESO price given by the solution $\tau_\star$ of the optimal stopping problem $(-Y, Q^\star)$:

$$\mathbb{E}_{Q^\star}\{e^{-r\,\tau_\star}\, X_{\tau_\star}\} \;\le\; \mathbb{E}_Q\{e^{-r\,(\tau\wedge T)}\, X_{\tau\wedge T}\}\,, \quad \text{for all } Q \in \mathcal{Q}\,. \tag{4.16}$$

(10) The optimal stopping times $\tau^\star$ and $\tau_\star$ are stopping times w.r.t. the filtration $(\mathcal{G}_t)_{0 \le t \le T}$. Thus, the information given by the takeover time τ is not necessary for establishing the no-arbitrage bounds for the ESO price. We point out that for the computation of the no-arbitrage bounds we can choose any EMM $Q \in \mathcal{Q}$, since $Y_{\tau^\star}$ and $-Y_{\tau_\star}$ are $\mathcal{G}_T$-measurable rv's, and on $\mathcal{G}_T$, all EMM's coincide, see Theorem 4.5. For simplicity, we choose $Q^\star$ for describing the no-arbitrage bounds.
(11) The pay-off profile process X is often discontinuous in T. At that time, the employee's reward for a takeover during $[0, T[$ is reduced to the plain ESO that matures in T. However,

by applying standard argumentation we can approximate X by a sequence of continuous pay-off processes (X_n) that converge to X, and the result of Proposition 4.9 still holds.
(12) We also mention, that possible discontinuities of the pay-off process X on $[0, T[$ can be accommodated in an additional cumulative income process extending the American option/optimal stopping problem. Furthermore, a hedging strategy corresponding to the upper no-arbitrage bound exists, if $\mathbb{E}_{Q^\star}\{\sup_{0 \le t \le T} X_t\} < \infty$, see Karatzas and Shreve (1998).

The remark on non-arbitrage bounds for the ESO prices is useful since American options are well-studied for financial markets driven by diffusions. In addition, the optimal stopping problem can be controlled quite easily in particular cases, cf. the American style call option on stock paying no dividends.

Corollary 4.10 *In the present setting, denote $h(x_1, x_2, t)$ the price function of the option granting the maximum of S^1 and S^2 with maturity t. The upper price bound of the ESO specified in Equation (4.11) is given by price of the American style option with possible pay off $\max\{c, h(T - t, S^1_t, S^2_t)\}$, for $t \in [0, T^V]$.*

Corollary 4.10 is a direct consequence of Proposition 4.9 since $\max\{S^1_T, S^2_T\} \ge (S^1_T - K)^+$. And moreover, exercising the American call on the maximum of S^1 and S^2 before T is not optimal, because early exercise in τ is subject to the inequality

$$S^1_T \, \mathbf{1}_{\{S^1_\tau \ge S^2_\tau\}} + S^2_T \, \mathbf{1}_{\{S^1_\tau < S^2_\tau\}} \le \max\{S^1_T, S^2_T\} \, .$$

Early exercise in $[0, T^V]$ may be favorable. The payment of c has maximal present value for $t = 0$, in $t = T^V$ the present value is just $c\, e^{-r\, T^V}$. Accordingly, not exercising in $t \in [0, T^V]$ should be at least worth the increment of the continuous payment stream $r\, c\, dt$. Early exercise can be ruled out by replacing c by $cS^0_t = c\, e^{r\, t}$. And hence, no costs incur when holding the option until T^V.

Corollary 4.11 *In the situation of Corollary 4.10, the ESO given in Equation (4.11) is modified by replacing c by $c\, S^0_t$. Then the upper price bound $\bar{\pi}$ of the modified ESO is*

$$\bar{\pi} = e^{-r\, T^V} \, \mathbb{E}_{Q^\star}\{\max\{c, h(S^1_{T^V}, S^2_{T^V}, T - T^V)\} \, .$$

Below, we will discuss how to incorporate both, early departure of the employee and takeover. The no-arbitrage bounds of the ESO will again be connected to the problem of optimal stopping, but in this case for the maximum of the takeover payoff and the payoff that is received by the employee when he leaves before maturity.

4.4.2 Performance Hurdles

The number of option granted to the employee is often linked to the performance of the company's share, see, e.g., the BHP case study in Maller et al. (2002). The option specification can also depend on the state of the market variables, share prices, at the vesting date T^V.

In the following, we study an ESO that grants $n(S_{T^V})$ call options in T^V, where n is a non-negative continuous function. The maturity of the option is $T > T^V$, and $T^\star = T - T^V$ is the time to maturity, when the option is granted to the employee. We consider an European style option with pay-off profile defined by f, where $f : \mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ is a continuous function, such that $f(S_{T^V}, \cdot)$ specifies the contract conditions in T^V and the pay-off is in T is $f(S_{T^V}, S_T)$.

Example. Suppose $d = 3$, and define a call option by

$$f(x, y) = (y_1 - K(x))^+ \ , \quad \text{for } x, y \in \mathbb{R}^4 \, ,$$

where the strike price $K(x) = x_1$ implies an at-the-money call option. The remuneration for a good performance of the company is incorporated in the number of share $n(x)$ that is a continuous function and increasing in x_1/x_2. By this specification, the value of the ESO increases if the share price S^1 of the company performs well in comparison to the market price of the reference asset S^2.

In this example, we explicitly address employee departure and takeover provisions simultaneously. The company that is a potential aspirant for a takeover is represented by its stock price process S^3. In case of departure before T and no takeover occurred prior to departure, the employee receives zero compensation. Whereas takeover before maturity T and prior to a possible departure T^D implies an extra compensation that is given by $\max\{S^1_{T^{TO}}, S^3_{T^{TO}}\}$, see previous section. Then the ESO is described by

$$X = \mathbf{1}_{\{\tau > T\}}\, n(S^1_{T^V}/S^2_{T^V})\, (S^1_T - S^1_{T^V})^+ + \mathbf{1}_{\{\tau \le T\}}\, \chi\, \bar{n} \max\{S^1_\tau, S^3_\tau\} \, , \tag{4.17}$$

where $\tau = T^D \wedge T^{TO}$, $\chi = \mathbf{1}_{\{T^{TO} \le T^D\}}$, and n is bounded by $\bar{n} > 0$.

By definition, τ is a stopping time, and χ is a $\mathcal{F}_\tau$-measurable random variable indicating whether $\tau = T^{TO}$ or not. The valuation of the ESO requires the information prior to τ and hence, we have the situation of one stopping time and the Cox property is preserved up to time τ under a change of measure. Thus, we may assume w.l.o.g. $Q \in \mathcal{Q}^{Cox}$. Further, note that the intensity λ of τ is given by $\lambda = \lambda^D + \lambda^{TO}$.

The price π_Q of the ESO given in Equation (4.17) has the representation

$$\begin{aligned} \pi_Q \;&=\; \mathbb{E}_{Q^\star}\left\{ n(S^1_{T^V}/S^2_{T^V})\, \exp\left(- \int_0^T (r + \lambda_u)\, du \right) (S^1_T - S^1_{T^V})^+ \right\} \\ &\quad + \bar{n} \int_0^T \mathbb{E}_{Q^\star}\left\{ \lambda^{TO}_t \exp\left(- \int_0^t (r + \lambda_u)\, du \right) \max\{S^1_t, S^3_t\} \right\} dt \, , \end{aligned}$$

Back in the general setting, we can find a strict upper price bound of an ESO that accommodates employee departure and takeover provisions. This is accomplished by extending Proposition 4.9 properly and of course applying standard argumentation as in Remark (9), we can also determine also a strict lower bound.

Proposition 4.12 *In the present setting, let an ESO be given by the processes* $X_t^D = F(t, S_t)$ *and* $X_t^{TO} = G(t, S_t)$*, and define* $\tau = T^D \wedge T^{TO}$*. The pay-off in* τ *is* X_τ *given by* $X_t = \mathbf{1}_{\{t \le \tau, T^D = \tau\}} X_\tau^D + \mathbf{1}_{\{t \le \tau, T^{TO} = \tau\}} X_\tau^{TO}$ *if* $\tau \le T$*, respectively by* $X_T = \max\{X_T^D, X_T^{TO}\}$ *if* $\tau > T$*, and* F *and* G *are measurable functions such that* X^D *and* X^{TO} *are continuous and bounded from below. Then*

$$\mathbb{E}_Q\{e^{-r\,(\tau \wedge T)} X_{\tau \wedge T}\} \le \mathbb{E}_{Q^\star}\{e^{-r\,\tau^\star} \max\{X^D, X^{TO}\}_{\tau^\star}\}\,, \quad \textit{for all } Q \in \mathcal{Q}\,, \tag{4.18}$$

where $\tau^\star$ *is the solution of the optimal stopping problem* $(Y, Q^\star)$*, and* Y *is given by* $Y_t = \max\{X_t^D, X_t^{TO}\}/S^0$*. Furthermore, the above equation gives a strict upper boundary for the ESO price.*

4.4.3 Random Vesting

Finally, we discuss the stochastics of the vesting time T^V. For example, the vesting of the ESO is connected to the outperformance of an a priori given reference index. The option is granted to the employee exactly when the stock price S^1 attains a pre-specified target value with respect to the benchmark/index given by S^2. Then, T^V turns out to be a first exit time, see Øksendal (1995), Ch. VII. Let us ignore departure of the employee and takeover in the beginning. Thus, the option is valued under the unique EMM Q for the complete sub-market $(S, (\mathcal{G}_t)_{0 \le t \le T})$, since by Theorem 4.5 the restriction of two arbitrary EMMs coincide on $\mathcal{G}_T$, i.e. $Q^1(A) = Q^2(A)$, for $A \in \mathcal{G}_T$, and $Q^1, Q^2 \in \mathcal{Q}$.

Let $U \subset \mathbb{R}^{d+1}$ be an open set, and define the vesting time T^V by

$$T^V = \inf\{t > 0 : S_t \notin U\}\,. \tag{4.19}$$

Furthermore, the ESO granted in T^V is assumed to be an European style option given by a measurable function $f \ge 0$ on $\mathbb{R}^{d+1}$ with maturity $T^V + T^\star$. Define the hitting distribution/harmonic measure μ_U^x according to Øksendal (1995), Ch. VII., p. 111, by

$$\mu_U^x(A) = Q^x\left(S_{T^V} \in A\right)\,, \text{ for } A \subset \partial U,\, x \in U,$$

and the value function of the contingent claim associated with f by

$$v(x) = \mathbb{E}_Q^x\{e^{-rT^\star} f(S^1_{T^\star})\}\,, \text{ for } x \in U.$$

Then the price π_Q of the ESO has the representation

$$\pi_Q(x) = \mathbb{E}_Q\left\{e^{-r(T^V + T^\star)} f(S_{T^V + T^\star})\right\} = \mathbb{E}_Q\left\{\frac{e^{-r\,T^\star} f(S_{T^V + T^\star})}{S^0_{T^V}}\right\} = \int_{\partial U} \frac{v(y)}{y_0}\, d\mu_U^x(y)\,, \tag{4.20}$$

where $x \in U$ is the initial value of the price process, i.e. $S_0 = x$.

Example. Let $d = 2$, and interpret S^2 as the price process of a reference index, e.g., Euro Stoxx 50 or S&P 500. The vesting date is defined by

$$T^V = \inf\{t > 0 : \ln(S_t^1/S_0^1) - \ln(S_t^2/S_0^2) \ge \kappa\}\,,$$

i.e. the time when the return of S^1 exceeds the return of the reference index S^2 by κ. Hence, the set U is defined by

$$U = \{(s_0, s_1, s_2) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ : \ln s_1 - \ln s_2 \leq \kappa\},$$

where the initial value for our problem are given by $S_0 = x = (1, 1, 1)$. And the boundary of U is of the form

$$\partial U = \mathbb{R}^+ \times \mathbb{R}^+ (e^\kappa, 1).$$

The option granted in T^V is an at-the-money call on the underlying S^1 with time to maturity equal to $T^\star$. Thus, the value function v does not depend on the state of the process in T^V, but is exactly the Black&Scholes price of an European at-the-money call, and hence constant $v(x) = v_{\sigma,T^\star}$. Then the price π of the specified ESO satisfies

$$\pi_Q = \pi_Q(1,1,1) = v_{\sigma,T^\star} \int\limits_{\partial U} \frac{1}{y_0}\, d\mu_U^{(1,1,1)}(y).$$

Observe that the integrand is a function of the first component y_0 and hence, we can simplify the above expression

$$\pi_Q = v_{\sigma,T^\star} \int\limits_{\mathbb{R}^+} \frac{1}{y_0}\, d\tilde{\mu}(y_0) = v_{\sigma,T^\star} \int\limits_{\mathbb{R}^+} e^{-r\,t}\, d\widetilde{Q}(T^V \leq t),$$

where $\tilde{\mu}(A_0) = \widetilde{Q}(S^0_{T^V} \in A_0) = \widetilde{Q}(e^{r\,T^V} \in A_0)$, $A_0 \subset \mathbb{R}^+$, and $\widetilde{Q} = Q^{(1,1,1)}$. Moreover, it can be shown that the first exit time T^V can be written as the time when the maximum of a Brownian motion with drift exceeds a specific level.

Define $Y_t = \ln(S^1_t/S^1_0) - \ln(S^2_t/S^2_0)$ for $t \geq 0$, then $Y \stackrel{d}{=} (a\,t + c\,B_t)_{t\geq 0}$, where B is an SBM, and a and c are depending on the parameters r and Σ determining the Q-dynamics of S. Let M denote the maximum of Y, i.e. $M = \sup_{0\leq t\leq \cdot} Y_t$, then we find

$$\widetilde{Q}(T^V \leq t) = \widetilde{Q}(M_t \leq \kappa).$$

The process M is a martingale iff $a = 0$ what is equivalent to $\|\sigma^1\| = \|\sigma^2\|$, i.e. the stock price S^1 and the benchmark S^2 have identical volatility. In this case the latter probability is well-known and given by

$$\widetilde{Q}(M_t \leq \kappa) = 2\widetilde{Q}\left(B_t \leq \frac{\kappa}{c}\right) - 1 = 2\Phi\left(\frac{\kappa}{c\sqrt{t}}\right) - 1, \quad \text{for } t > 0,$$

by the reflection principle for Brownian motion, see Protter (1990), Ch. I, Theorem 33, where Φ is the distribution function of a standard normal rv.

At the end of this consideration, we discuss how to incorporate performance linked vesting and employee departure and/or takeover provisions. Suppose that the departure of the employee forfeits the ESO, even after the option is granted. Then ESO valuation becomes very much related to pricing of credit derivatives. Especially, for $Q \in \mathcal{Q}^{exp}$ Equation (4.20) extends to

$$\pi_Q(x) = \mathbb{E}_Q \left\{ \mathbf{1}_{\{T^V + T^\star < T^D\}} e^{-r(T^V + T^\star)} f(S_{T^V + T^\star}) \right\} = \int_{\partial \bar{U}} \frac{\bar{v}(y)}{y_0} \, d\bar{\mu}^x_{\bar{U}}(y) \,, \tag{4.21}$$

where we use the Cox property, i.e. by conditioning on $\mathcal{G}_\infty$ we find

$$\mathbb{E}_Q \left\{ \mathbf{1}_{\{T^V + T^\star < T^D\}} e^{-r(T^V + T^\star)} f(S_{T^V + T^\star}) \right\} = \mathbb{E}_Q \left\{ e^{-\int_0^{T^V + T^\star} (r + \lambda^D_u)\, du} f(S_{T^V + T^\star}) \right\} ,$$

and define $\bar{v}$ by

$$\bar{v}(x) = \mathbb{E}^x_Q \{ f(S^1_{T^\star}) / \bar{S}^0_{T^\star} \} \,, \text{ for } x \in \bar{U}, \text{ and } \bar{S}^0 = S^0 \, e^{\int_0^\cdot \lambda^D_u \, du} \,,$$

and the hitting distribution

$$\mu^x_{\bar{U}}(A) = Q^x \left(\bar{S}_{T^V} \in A \right) \,, \text{ for } A \subset \partial \bar{U}, \; x \in \bar{U}, \text{ and } \bar{S} = (\bar{S}^0, S^1, ..., S^d) \,.$$

To describe the set $\bar{U}$ we make use of the fact that $Q \in \mathcal{Q}^{exp}$, i.e. $\lambda^D_t(\omega) = \lambda$, for some non-negative constant λ. And $\bar{U}$ is defined that the vesting time T^V satisfies Equation (4.19) when replacing S and U by $\bar{S}$ and $\bar{U}$.

$$\bar{U} = \left\{ (u_0^{\frac{r+\lambda}{r}}, u_1, ..., u^d) : (u_0, ..., u_d) \in U \right\} .$$

The valuation formula in Equation (4.21) holds in more general setting. If the departure intensity λ^D is a diffusion process, i.e. $\lambda_t = h(t, W_t)$, then the presented methodology can also be applied for $Q \in \mathcal{Q}^{Cox}$ by extending the state space properly.

4.5 Effects of Inside Information

In the following we give some examples that illustrate Assumption 4.3, and emphasize the necessity of the (H) hypothesis that ensures an arbitrage-free setting of the proposed reduced form model.

4.5.1 Brownian Random Time

We briefly summarize the example given by Blanchet-Scalliet and Jeanblanc (2002): The stopping time T^D is given by the time when S^1 crosses a given boundary $x > 0$ in downward direction for the last time until the end of the time horizon T.

$$T^D = \sup\{0 \le t \le T : S^1_t \ge x\} \,.$$

Arbitrage opportunities are implied by this market specification. The stopping time T^D announces when S^1 stays below the given level x, and hence the martingale property of

W is lost on the enlarged filtration $(\mathcal{F}_t)_{0\le t\le T}$. Obviously, the (H) hypothesis that is part of Assumption 4.3 (a) does not hold in this setting.

To clarify the structure of the model specification of the example by Blanchet-Scalliet and Jeanblanc (2002), observe that the market model $(S,(\mathcal{G}_t)_{0\le t\le T})$ is complete and hence arbitrage-free, whereas the market model $(S,(\mathcal{F}_t)_{0\le t\le T})$ bears arbitrage opportunities.

4.5.2 Brownian Bridge Specification

Given the reduced form model definition of Chapter 2, we assume $d=1$, i.e. we have one risky asset S^1, and the riskless bank account given by S^0. For the remainder of this example, we will emphasize on $\tau = T^D$ with indicator process $N = \mathbf{1}_{\{\,.\le\tau\}}$, and we ignore the random time T^{TO}.
First, we construct the example mathematically. Let the filtration $(\mathcal{C}_t)_{0\le t\le T}$ be given by $\mathcal{C}_t = \mathcal{G}_T \vee \mathcal{F}^N_t$, where $(\mathcal{F}^N_t)_{0\le t\le T}$ is the natural filtration of N. Suppose $(\mathcal{C}_t)_{0\le t\le T}$ satisfies the usual conditions, and N admits a predictable $(\mathcal{C}_t)_{0\le t\le T}$-intensity λ, i.e.

$$M = N - \int_0^{\cdot\wedge\tau} \lambda_t\, dt\,,$$

is a $(\mathcal{C}_t)_{0\le t\le T}$-martingale. The form of λ_t is relevant on the set $\{\tau > t\}$, and following Blanchet-Scalliet and Jeanblanc (2002) (comment afterwards Corollary 1), we find that λ_t is equal to a $\mathcal{G}_T$-measurable rv on $\{\tau > t\}$. Thus we can assume w.l.o.g. λ_t is $\mathcal{G}_T$-measurable, for all $t\in[0,T]$. This description of the model coincides with the first jump of a Cox process on the filtration $(\mathcal{C}_t)_{0\le t\le T}$, see Brémaud (1981), Chap. II, and we find

$$P\left(\tau > t \mid \mathcal{G}_T\right) = e^{-\int_0^t \lambda_u\, du}\,, \quad \text{for } 0\le t\le T,$$

what is just Equation (4.1). But in this situation, λ_t is not necessarily $\mathcal{G}_t$-measurable.
The Brownian bridge specification is now formalized by defining

$$\lambda_t = f(S^1_T)\,, \quad \text{for } 0\le t\le T,$$

where $f\in C^2(\mathbb{R}^+,\mathbb{R}^+)$ is a continuous and bounded function.

The intensity $\hat\lambda$ of N for the history $(\mathcal{F}_t)_{0\le t\le T}$ generated by W and N is given by Brémaud (1981), Theorem 14, Ch. II.

$$\hat\lambda_t = \mathbb{E}_P\left\{\lambda_t \mid \mathcal{F}_t\right\} = \mathbb{E}_P\left\{f(S^1_T)\mid\mathcal{F}_t\right\} = \mathbb{E}_P\left\{f(S^1_T)\mid\mathcal{G}_t\right\}, \quad \text{on } \{\tau>t\},\ \text{for } 0\le t\le T,$$

where we apply the same argument as in the above situation to replace $\mathcal{F}_t$ by $\mathcal{G}_t$.

We briefly address the representation of $\hat\lambda$ as solution of a partial differential equation applying the Feynman-Kac framework, see Karatzas and Shreve (1988), Ch. 5.7. Denote A the infinitesimal generator of the (homogeneous) diffusion S^1, then

$$\hat\lambda_t = v(t, S^1_t) \quad \text{for } 0\le t\le T,$$

where $v(t,x)$ is the solution of the partial differential equation

$$-\frac{\partial v}{\partial t} = Av\,, \quad \text{and} \quad v(T,\cdot) = f\,, \quad \text{and} \quad A = \mu x \frac{\partial}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2}\,.$$

Using once more the infinitesimal generator A, we can express $\hat{\lambda}$ as sum of a known part and the average/expectation of the future random development.

$$\hat{\lambda}_t = f(S^1_t) + \mathbb{E}_P\left\{ \int_t^T (Af)(S^1_u)\,du \,\middle|\, \mathcal{G}_t \right\}\,, \quad \text{for } 0 \le t \le T.$$

Let us summarize the construction, so far. The model contains a process W that describes the evolution of the stock price process S^1, and W is a Brownian motion in its natural filtration $(\mathcal{G}_t)_{0\le t\le T}$. A random time τ is added with indicator process N. Conditioned on the complete market information $\mathcal{G}_T$, τ is a stopping time w.r.t. to this filtration with $(\mathcal{C}_t)_{0\le t\le T}$-intensity $\lambda_t = f(S^1_T)$. This specification has three major consequences:

(a) The process N admits the $(\mathcal{G}_t)_{0\le t\le T}$-predictable intensity $\hat{\lambda}$ on $(\mathcal{F}_t)_{0\le t\le T}$, and hence Assumption 4.3 (c) holds.

(b) The process N carries the first jump of a Cox process with $(\mathcal{C}_t)_{0\le t\le T}$-intensity $\lambda \neq \hat{\lambda}$.

(c) The (H) hypothesis is not fulfilled for the Brownian bridge specification.

The model is constructed such that (a) holds, and we already verified (b) by using a result of Brémaud (1981), what may surprising in some way, but this is due to introducing $(\mathcal{C}_t)_{0\le t\le T}$, and a rigorous interpretation of the definition of a Cox process. In the following, consequence (c) is shown, and as well the terminology *Brownian bridge specification* is explained in detail.
In the second stage, we discuss the economic interpretation of the given model. The employee is the insider in this market specification. He knows the stock price at the end of the time horizon in advance, so his information/filtration is given by $(\widetilde{\mathcal{G}}_t)_{0\le t\le T}$, where $\widetilde{\mathcal{G}}_t = \mathcal{G}_t \vee \sigma(W_T)$, for $0 \le t \le T$. Note, that the sigma fields generated by W_T and S^1_T coincide, $\sigma(W_T) = \sigma(S_T)$, since the constant coefficient setting implies $S^1_T = S^1_0 \exp((\mu - \sigma^2/2)\,T + \sigma W_T)$. On the enlarged (insider) filtration $(\widetilde{\mathcal{G}}_t)_{0\le t\le T}$ the process W is no longer a Brownian motion but still a special semimartingale with Doob-Meyer decomposition of the following form

$$W_t = \widetilde{W}_t + \int_0^t \frac{W_T - W_s}{T - s}\,ds\,, \quad \text{for } 0 \le t \le T, \tag{4.22}$$

where $\widetilde{W}$ is a Brownian motion w.r.t. $(\widetilde{\mathcal{G}}_t)_{0\le t\le T}$ that is independent of W_T. This representation is also known as *Brownian bridge*, see for example Jeulin and Yor (1979). Such specifications for insider markets are well studied, see Kyle (1985) or Back (1992).

Here we have a different situation: The insider does not trade in the market (usually employees are allowed only limited trading in the stock of their employing company), but the inside information enters the market filtration by the event of employee's departure from the company, i.e. when enlarging from $(\mathcal{G}_t)_{0\leq t\leq T}$ to $(\mathcal{F}_t)_{0\leq t\leq T}$. So, the insider information is partially revealed when τ occurs. More precisely, it can be shown that the stopped Brownian motion W^τ remains a martingale on $(\mathcal{F}_t)_{0\leq t\leq T}$ by applying results of Elliot et al. (2000), Section 3. After τ the martingale property of W is lost since the (H) hypothesis does not hold, and this implies a non-trivial Doob-Meyer decomposition of W after τ. (Unfortunately, we can not calculated this decomposition in an explicit form.)

Finally, we show that the (H) hypothesis in Assumption 4.3 (a) does not hold for the Brownian bridge specification. This is obvious by the following result of Elliot et al. (2000): The (H) hypothesis is equivalent to

$$P(\tau \leq s \,|\, \mathcal{G}_t) = P(\tau \leq s \,|\, \mathcal{G}_T)\,, \quad \text{for } 0 \leq s \leq t \leq T.$$

Now, suppose the (H) hypothesis holds, we find for $s = t$

$$P(\tau > t \,|\, \mathcal{G}_t) = \exp\left(-t\, f(S_T^1)\right),$$

and $f(S_T)$ is $\mathcal{G}_t$-measurable rv, for all $0 < t \leq T$. This is valid only if f is a constant function (pathological case), and we obtain the desired contradiction for the (H) hypothesis.

Remark. (13) In the credit risk literature it is often assumed that working under the (H) hypothesis is equivalent to Cox process modelling, see, e.g., Blanchet-Scalliet and Jeanblanc (2002), Section 5. The above example demonstrates that this does not hold in general. The reason for this contradiction is the specific understanding of the term Cox process in the credit risk community: Implicitly it is assumed that Cox process modelling implies the existence of a *state process* W *driving* the intensity λ, and by defining a *minimal*/natural filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ such that the Cox process and the driving state process are both adapted to this filtration. And this filtration is studied in credit risk literature.

The above definition of a Cox process is a restrictive interpretation of the general definition, see Brémaud (1981). The intensity λ needs to be measurable w.r.t. to a given sigma field. In the above example, we choose $\mathcal{C}_0 = \mathcal{G}_T \supset \sigma(W_T)$, and define the intensity by the rv W_T. And hence W as process does *not drive* the Cox process N. Finally note, $(\mathcal{C}_t)_{0\leq t\leq T}$ is in some sense the *maximal* filtration for N, since it is the largest filtration such that $N - \int_0^\cdot \lambda_t\, dt$ remains a martingale.

Thus, we have constructed an example with a Cox process where the (H) hypothesis does not hold. Nevertheless, the converse still holds: the (H) hypothesis implies Cox process modelling (in its restricted form), see Lemma 2 in Blanchet-Scalliet and Jeanblanc (2002).

4.6 Proofs, Definitions, and useful Results

For the definition of strong orthogonality for square integrable martingales we refer to Protter (1990), Ch. IV.

Definition 4.13 *Denote* $\mathbf{M}^2$ *the set of all square integrable martingales with initial value zero in* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, P)$. *Two martingales* $M, N \in \mathbf{M}^2$ *are said to be* **strongly orthogonal** *if their product* $L = MN$ *is a (uniformly integrable) martingale.*

By the Kunita-Watanabe inequality we find a formulation of strong orthogonality in terms of the quadratic covariation process. Two martingales $M, N \in \mathbf{M}^2$ *are strongly orthogonal if and only if* $[M, N]$ *is a uniformly integrable martingale.*

Example. Let M and N be two martingales in $\mathbf{M}^2$.
(a) Suppose M has continuous paths and, the paths of N are of finite variation, then $[M, N] = 0$, implying that M, N are strongly orthogonal. Especially, if M is a Brownian motion and N is a compensated point process the strong orthogonality holds.
(b) Suppose M and N are compensated point processes both admitting intensities, then $[M, N] = \sum_{0<t\leq \cdot} \Delta M_t \Delta N_t$ collects all common jumps, and is an increasing process of finite variation. Thus, two point processes are strongly orthogonal if and only if they have no common jumps.

We define the predictable representation property along the lines of Protter (1990), Ch. IV:

Definition 4.14 *Denote* $\mathbf{M}^2$ *the set of all square integrable martingales with initial value zero in* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, P)$, *and* $\mathcal{A}$ *a finite set of martingales in* $\mathbf{M}^2$. $\mathcal{A}$ *has the* **predictable representation property**, *if every (square integrable) martingale* $M \in \mathbf{M}^2$ *can be represented as stochastic integral with respect to the elements of* $\mathcal{A}$:

$$\mathbf{M}^2 = \left\{ M : M = \sum_i \int H^i \, dM^i, \; M^i \in \mathcal{A} \right\} \tag{4.23}$$

and each H^i *is predictable such that*

$$\mathbb{E}_P \left\{ \int_0^T (H_t^i)^2 \, d[M^i, M^i]_t \right\} < \infty \,. \tag{4.24}$$

For the reduced form model proposed here the predictable representation property holds with respect to the Brownian motion W, and the martingales M^D and M^{TO}.

Theorem 4.15 (Kusuoka, 1999) *Let the reduced form model be given by Definition 4.1, and Definition 4.2, and Assumption 4.3. The set* $\mathcal{A} = \{W^1, ..., W^d, M^D, M^{TO}\}$ *has the predictable representation property: Every square integrable* P*-martingale* M *for* $(\mathcal{F}_t)_{0 \leq t \leq T}$ *has a representation*

$$M_t = M_0 + \sum_{i=1}^d \int_0^t H_s^i \, dW_s^i + \int_0^t K_s^D \, dM_s^D + \int_0^t K_s^{TO} \, dM_s^{TO} \,, \quad \textit{for } 0 \leq t \leq T, \tag{4.25}$$

for predictable process $H = (H^1, ..., H^d)$, and K^D, K^{TO} that satisfy

$$\mathbb{E}_P\left\{\int_0^T \|H_t\|^2 dt\right\} < \infty,\ \mathbb{E}_P\left\{\int_0^T (K_t^D)^2 \lambda_t^D\, dt\right\} < \infty,\ \mathbb{E}_P\left\{\int_0^T (K_t^{TO})^2 \lambda_t^{TO}\, dt\right\} < \infty. \tag{4.26}$$

Furthermore, if $\lambda^D, \lambda^{TO} > 0$ this representation is unique.

Proof of Theorem 4.5. The measure $Q \sim P$ is a martingale measure, if the discounted stock prices $S^1/S^0, ..., S^d/S^0$ are martingales under this specific measure Q, see Harrison and Pliska (1981). Any arbitrary equivalent measure $Q \sim P$ can be characterized by the predictable process (ψ, ϕ), and combining Definition 4.1 and Theorem 4.4 the Q-dynamics of $S^1/S^0, ..., S^d/S^0$ are

$$d(S^k/S^0)_t = S_t^k/S_t^0 \left((\mu^k - r)\, dt + \sigma^k\, \psi_t^\top\, dt + \sigma^k\, d\widetilde{W}_t \right), \quad \text{for } 0 \le t \le T \text{ and } k = 1, ...d,$$

where $\widetilde{W}$ a Q-SBM. Thus the discounted stock prices $S^1/S^0, ..., S^d/S^0$ are Q-martingales iff Equation (4.8) holds

$$\psi_t^\top = -\Sigma^{-1}(\mu - r\,\mathbf{e}), \quad \text{for } 0 \le t \le T.$$

This is exactly the classical result for the (complete) Black&Scholes market discussed by Harrison and Pliska (1981), except for the predictable process ϕ that is not affected by the martingale restrictions, and hence arbitrary, of course subject to certain regularity conditions. □

In order to prove Theorem 4.8 we need the following lemma.

Lemma 4.16 *On a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le T}, P)$ let $(\mathcal{G}_t)_{0 \le t \le T}$ be filtration satisfying the usual conditions with $\mathcal{G}_t \subset \mathcal{F}_t$, and define $\mathbf{M}^2_{\mathcal{G}} = \{M \in \mathbf{M}^2 : M \text{ is } (\mathcal{G}_t)_{0 \le t \le T}\text{-adapted}\}$. Suppose $M \in \mathbf{M}^2$ is strongly orthogonal to $\mathbf{M}^2_{\mathcal{G}}$, then the projection $\hat{M}$ of M on $\mathbf{M}^2_{\mathcal{G}}$ is constant, i.e.*

$$\hat{M}_t = \mathbb{E}_P\left\{M_T \,|\, \mathcal{G}_t\right\} = 0\,, \quad \textit{for } 0 \le t \le T. \tag{4.27}$$

Proof. By definition $\hat{M}$ is a $(\mathcal{G}_t)_{0 \le t \le T}$-adapted martingale, hence $\hat{M} \in \mathbf{M}^2_{\mathcal{G}}$ is strongly orthogonal to M as asserted. By elementary calculations we find

$$\mathbb{E}_P\left\{\hat{M}_T\, M_T\right\} = \mathbb{E}_P\left\{\mathbb{E}_P\left\{\hat{M}_T\, M_T \,|\, \mathcal{G}_T\right\}\right\} = \mathbb{E}_P\left\{\hat{M}_T\, \mathbb{E}_P\left\{M_T \,|\, \mathcal{G}_T\right\}\right\} = \mathbb{E}_P\left\{\hat{M}_T^2\right\},$$

and also $\mathbb{E}_P\left\{\hat{M}_T\, M_T\right\} = 0$, because of the strong orthogonality. We conclude $\hat{M}_T = 0$ almost surely, and also $\hat{M}_t = \mathbb{E}_P\left\{M_T \,|\, \mathcal{G}_t\right\} = 0$ almost surely, for $0 \le t \le T$. □

Example. Lemma 4.16 has an immediate consequence for the reduced form model given by Definition 4.1, Definition 4.2, and Assumption 4.3. Let $\mathcal{A} = \{W^1, ..., W^d\}$, then it is a well-known fact that $\mathcal{A}$ has the predictable representation property w.r.t. $(\mathcal{G}_t)_{0\le t\le T}$, i.e. $\mathcal{S}_{\mathcal{G}}(W) = \mathbf{M}^2_{\mathcal{G}}$. And by applying the (H) hypothesis we obtain $\mathbf{M}^2_{\mathcal{G}} = \mathcal{S}_{\mathcal{G}}(W) \subset \mathcal{S}(W)$ Define the (local) $(\mathcal{F}_t)_{0\le t\le T}$-martingale Y by

$$Y = \mathcal{E}(-M^D) = (1 - N^D)\exp\left(\int_0^\cdot \lambda_u^D\, du\right),$$

where $\mathcal{E}$ denotes the stochastic exponential. W.l.o.g. assume Y is square integrable and hence a martingale in M^2. Now observe, Y is strongly orthogonal to W and also strongly orthogonal to $\mathbf{M}^2_{\mathcal{G}} \subset \mathcal{S}(W)$. For each t, the process stopped in t is Y^t and Y^t is again strongly orthogonal to $\mathbf{M}^2_{\mathcal{G}}$, and $\mathbb{E}_P\{Y_T^t \,|\, \mathcal{G}_T\} = 1$ by Lemma 4.16, what implies

$$P(T^D > t \,|\, \mathcal{G}_T) = \mathbb{E}_P\{(1 - N_{t\wedge T}) \,|\, \mathcal{G}_T\} = \exp\left(-\int_0^t \lambda_u^D\, du\right), \quad \text{for } 0 \le t \le T.$$

This relation characterizes the Cox process and is given in Equation (4.1) as motivation.

Proof of Theorem 4.8. For proving that $Q^\star$ minimizes the $L^2(P)$-distance $D(\cdot, P)$ and the relative entropy $H(\cdot|P)$ over all $Q \in \mathcal{Q}$, we use the structure implied by the (H) hypothesis. Starting with the enlarged filtration $(\mathcal{F}_t)_{0\le t\le T}$, we shrink the information sets to $(\mathcal{G}_t)_{0\le t\le T}$, and apply that $H(\cdot|P)$ and $D(\cdot, P)$ are increasing in the sigma field.

Recall, if Q and P are probability measures on $(\Omega, \mathcal{F})$ and $\mathcal{G} \subset \mathcal{F}$ is a sigma field, then define the probability measure $Q_{\mathcal{G}}$ by the restriction of Q on $\mathcal{G}$ and

$$0 \;\le\; D(Q_{\mathcal{G}}, P) \;\le\; D(Q, P), \text{ and} \tag{4.28}$$

$$0 \;\le\; H(Q_{\mathcal{G}}|P) \;\le\; H(Q|P). \tag{4.29}$$

Note that $Q_{\mathcal{G}}$ is given by its Radon-Nikodym derivative $L_{\mathcal{G}}$ w.r.t. P by $L_{\mathcal{G}} = \mathbb{E}_P\{L|\mathcal{G}\}$, where $L = dQ/dP$. Furthermore, Equations (4.28–4.29) are direct consequences of Jensen's inequality applied to L for the convex functions $\varphi_D(x) = x^2$ and $\varphi_H(x) = x \ln x$, and the sigma field $\mathcal{G}$, of course under the measure P.

First, observe that the density process $L^\star$ of the EMM $Q^\star$ w.r.t. P is $(\mathcal{G}_t)_{0\le t\le T}$-adapted, see Theorem 4.4. By virtue of the (H) hypothesis, the Doob-Meyer decomposition of S is identical on both filtrations, $(\mathcal{G}_t)_{0\le t\le T}$ and $(\mathcal{F}_t)_{0\le t\le T}$. Thus, $L^\star$ is the density process of an EMM for the financial market $(S, (\mathcal{G}_t)_{0\le t\le T})$. Furthermore, this market is complete and hence $Q^\star_{\mathcal{G}_T}$ is the unique EMM, where $Q^\star_{\mathcal{G}_T}$ is the restriction of $Q^\star$ on $\mathcal{G}_T$. Moreover, variance and relative entropy are functionals of the Radon-Nikodym derivative $L^\star$, and hence $D(Q^\star, P) = D(Q^\star_{\mathcal{G}_T}, P)$, and $H(Q^\star|P) = H(Q^\star_{\mathcal{G}_T}|P)$.

Now, let $Q \in \mathcal{Q}$ be an arbitrary equivalent martingale measure, and denote $X = S/S^0$ the discounted price process. Then X is also a $(Q, (\mathcal{G}_t)_{0\le t\le T})$-martingale, since X is adapted to $(\mathcal{G}_t)_{0\le t\le T}$. By definition, the restricted measure $Q_{\mathcal{G}_T}$ is identical to Q on $\mathcal{G}_T$, and hence X is a $(Q_{\mathcal{G}_T}, (\mathcal{G}_t)_{0\le t\le T})$-martingale. But the financial market $(S, (\mathcal{G}_t)_{0\le t\le T})$ is complete with unique martingale measure $Q^\star_{\mathcal{G}_T}$ what implies $Q_{\mathcal{G}_T} = Q^\star_{\mathcal{G}_T}$.

Applying Equation (4.28), we find

$$D(Q^\star, P) = D(Q^\star_{\mathcal{G}_T}, P) = D(Q_{\mathcal{G}_T}, P) \le D(Q, P).$$

In the same manner we conclude $H(Q^\star|P) \leq H(Q|P)$ using Equation (4.29).

It remains to prove the additive decomposition of the relative entropy given in Equation (4.10). Let Q^ϕ be a measure in $\mathcal{Q}$. The Radon-Nikodym derivative $L_T^\phi = dQ^\phi/dP$ is given by the pair $(\psi^\star, \phi)$, see Theorem 4.4, and allows a factorization because of the special form of strong orthogonality between W and $M = (M^D, M^{TO})$, i.e. $[W, M] = 0$:

$$L^\phi = \mathcal{E}(\int \psi^\star \, dW + \int (\phi - 1) \, dM) = \mathcal{E}(\int \psi^\star \, dW) \, \mathcal{E}(\int (\phi - 1) \, dM) \,, \tag{4.30}$$

where $\mathcal{E}$ denotes the stochastic exponential. The measure $Q^\star$ is given by the density process $L^\star = L^{(1,1)}$ with respect to P. We find

$$\begin{aligned} H(Q^\phi|P) &= \int \log \left(\frac{dQ^\phi}{dQ^\star} \frac{dQ^\star}{dP} \right) \, dQ^\phi \\ &= \int \log \left(\frac{dQ^\phi}{dQ^\star} \right) \, dQ^\phi + \int \log \left(\frac{dQ^\star}{dP} \right) \, dQ^\phi \\ &= H(Q^\phi|Q^\star) + \int \log \left(\frac{dQ^\star}{dP} \right) \, dQ^\star + \int \log \left(\frac{dQ^\star}{dP} \right) \left(\frac{dQ^\phi}{dQ^\star} - 1 \right) \, dQ^\star \\ &= H(Q^\phi|Q^\star) + H(Q^\star|P) + \Delta_P(Q^\phi|Q^\star) \,, \end{aligned}$$

where

$$\Delta_P(Q^\phi|Q^\star) = \int \log \left(\frac{dQ^\star}{dP} \right) \left(\frac{dQ^\phi}{dQ^\star} - 1 \right) \, dQ^\star \,.$$

The expression $\Delta_P(Q^\phi|Q^\star)$ is an expectation w.r.t. the measure $Q^\star$. Applying Equation (4.30) and the iterated expectation conditioning on $\mathcal{G}_T$ results in

$$\Delta_P(Q^\phi|Q^\star) = \mathbb{E}_{Q^\star} \left\{ \log \left(\frac{dQ^\star}{dP} \right) \mathbb{E}_{Q^\star} \left\{ \left(\mathcal{E}(\int (\phi - 1) \, dM)_T - 1 \right) \Big| \mathcal{G}_T \right\} \right\}$$

since $L_T^\star = dQ^\star/dP$ is $\mathcal{G}_T$-measurable. Observe by Theorem 4.4 that M is a $Q^\star$-martingale and the (H) hypothesis is preserved under the change of measure, see Remark (6). The $Q^\star$-martingale $Y = \mathcal{E}(\int (\phi - 1) \, dM)_T - 1$ is strongly orthogonal to $W^\star = W - \int \psi^\star dt$, since M is strongly orthogonal to $W^\star$, and applying Lemma 4.16 gives us

$$\mathbb{E}_{Q^\star} \left\{ \left(\mathcal{E}(\int (\phi - 1) \, dM)_T - 1 \right) \Big| \mathcal{G}_T \right\} = 0$$

Thus $\Delta_P(Q^\phi|Q^\star) = 0$, and this yields the identity $H(Q^\phi|P) = H(Q^\phi|Q^\star) + H(Q^\star|P)$. In the given situation Lemma 4.16 is applicable under the measure $Q^\star$ because of the following reasons: The (H) hypothesis guarantees the predictable representation property of $W^\star$ with respect to the set $\mathbf{M}^2_\mathcal{G}(Q^\star) = \{M \in \mathbf{M}^2(Q^\star) : M \text{ is } (\mathcal{G}_t)_{0 \leq t \leq T}\text{-adapted}\}$, and note by definition, $W^\star$ is $(\mathcal{G}_t)_{0 \leq t \leq T}$-adapted. This gives us $\mathbf{M}^2_\mathcal{G}(Q^\star) \subset \mathcal{S}(W^\star)$, and because of Lemma 2 in Protter, Ch. IV, Sec. 3, implying the strong orthogonality of Y to $\mathcal{S}(W^\star)$. □

Proof of Proposition 4.9. The proof is divided into two parts. First, we show that the optimal stopping time $\tau^\star$ provides an upper bound for the expected value $\mathbb{E}_Q\{Y_{\tau \wedge T}\}$ for

all EMM $Q \in \mathcal{Q}$, Part (a). Then we construct a sequence of EMMs $Q^n \in \mathcal{Q}$ such that $\mathbb{E}_{Q^n}\{Y_\tau\} \nearrow \mathbb{E}_{Q^\star}\{Y_{\tau^\star}\}$, for $n \to \infty$, and hence the upper bound is strict, Part (b).
(a) This part of the proof is based on the (H) hypothesis that holds in our setting, see Assumption 4.3 (a): The stopping time τ does not affect the martingale dynamics when enlarging $(\mathcal{G}_t)_{0\leq t\leq T}$ to $(\mathcal{F}_t)_{0\leq t\leq T}$. Thus, τ does not contain future information of the price processes and stopping in τ is effectively arbitrary, and this strategy is dominated by an optimal stopping rule.
We fix an arbitrary EMM $Q \in \mathcal{Q}$, and define the conditional survival probability L of τ given $\mathcal{G}_T$ by $L_t = Q(\tau > t \,|\, \mathcal{G}_T)$. According to Assumption 4.3, the stopping time τ admits a P-intensity λ^P that is $(\mathcal{G}_t)_{0\leq t\leq T}$-predictable. This structure is preserved when changing to the measure Q, see Theorem 4.5 and Remark (6), and hence τ admits a Q-intensity λ that is $(\mathcal{G}_t)_{0\leq t\leq T}$-predictable. Thus $L_t = \exp(-\int_0^t \lambda_u \, du)$, and:

$$\begin{aligned}
\mathbb{E}_Q\{Y_{\tau\wedge T}\} &= \mathbb{E}_Q\left\{\mathbb{E}_Q\left\{Y_{\tau\wedge T} \mid \mathcal{G}_T\right\}\right\} \\
&= \mathbb{E}_Q\left\{\int_0^T Y_t \, dQ(\tau \leq t \mid \mathcal{G}_T) + Y_T \, Q(\tau > T \mid \mathcal{G}_T)\right\} \\
&= \mathbb{E}_{Q^\star}\left\{\int_0^T Y_t \, d\overline{L}_t + Y_T \, L_T\right\},
\end{aligned}$$

where $\overline{L} = 1 - L$. Note, that last line of the equation the measure Q is replaced by the EMM $Q^\star$. This particular choice emphasizes that $\int_0^T Y_t \, d\overline{L}_t + Y_T \, L_T$ is $\mathcal{G}_T$-measurable, and on this σ-field all EMMs coincide. In the above representation the process L contains the structure of Q, since L is defined by the Q-intensity of τ. The expected value $\mathbb{E}_{Q^\star}\{Y_{\tau\wedge T}\}$ can be interpreted as taking the expectation after averaging Y_t over time with weighting scheme/density $\frac{d\overline{L}}{dt} = \lambda \, L$. This average value $\int_0^T Y_t \, d\overline{L}_t + Y_T \, L_T$ is suboptimal and can be dominated by an optimal stopping strategy given by $\tau^\star$.

$$\mathbb{E}_Q\{Y_{\tau\wedge T}\} \leq \mathbb{E}_{Q^\star}\{Y_{\tau^\star}\}, \quad \text{for all } Q \in \mathcal{Q},$$

where $\tau^\star$ is the solution to the optimal stopping problem $(Y, Q^\star)$.
(b) It remains to prove that the above established bound by the optimal stopping time $\tau^\star$ is a strict bound. Following Theorem 4.5, the intensity of τ is any arbitrary non-negative $(\mathcal{G}_t)_{0\leq t\leq T}$-predictable process λ. Let us now specify a sequence of EMMs $(Q^n)_{n\geq 1}$ by defining the associated intensity process λ^n

$$\lambda_t^n = n \, \mathbf{1}_{\{\tau^\star < t\}}, \quad \text{for } 0 \leq t \leq T, \text{ all } n \geq 1.$$

The process λ^n is left continuous and $(\mathcal{G}_t)_{0\leq t\leq T}$-adapted, hence predictable. By construction, we find $(\tau, W, Q^n) \xrightarrow{d} (\tau^\star, W, Q^\star)$, and by continuity of Y

$$\mathbb{E}_{Q^n}\{Y_\tau\} \nearrow \mathbb{E}_{Q^\star}\{Y_{\tau^\star}\}, \quad \text{for } n \to \infty,$$

and hence the upper boundary is strict. □

Proof of Proposition 4.12. We apply the same arguments as in the proof of Proposition 4.9 to the function $\max\{F, G\}$. This yields us the upper boundary Equation (4.18) given by the optimal stopping time $\tau^\star$. To show that the boundary is strict, we have to construct a sequence of EMMs $(Q^n)_{n\geq 1}$ such that the prices/expectations converge to the upper boundary. In contrast to Proposition 4.9, we have two stopping times T^D and T^{TO}. We extend the construction by defining the intensities by

$$\begin{aligned} \lambda_t^{D,n} &= n\,\mathbf{1}_{\{\tau^\star<t;\, X^D_{\tau^\star}\geq X^{TO}_{\tau^\star}\}}\,, \\ \lambda_t^{TO,n} &= n\,\mathbf{1}_{\{\tau^\star<t;\, X^{TO}_{\tau^\star}\geq X^{D}_{\tau^\star}\}}\,, \quad \text{for } 0\leq t\leq T,\ \text{all } n\geq 1\,. \end{aligned}$$

The stopping time $T^D \wedge T^{TO}$ converges to $\tau^\star$ and furthermore, the maximal pay-off is chosen by the above construction of the intensities. Thus, the claimed result follows, see proof of Proposition 4.9. □

Chapter 5

How to Explain a Corporate Credit Spread

5.1 Introduction

In recent years many models and ideas concerning credit risk went public. Here, we refer to three related approaches: The reduced-form model defines default as an unpredictable event that is governed by a hazard-rate process, among others see Duffie and Singleton (1999). Jarrow, Lando and Turnbull (1997) share the intensity based approach, but they focus on transitions inbetween different rating classes incorporating a homogeneous continuous time Markov chain with rating classes in a discrete state space. Lando (1997) presents a technique of adding a certain set of explaining variables to this model, thus the Markov chain becomes heterogeneous. A more global point of view is stated in Schönbucher (1998) where the term structure model of Heath, Jarrow and Morton (1992) is extended by an additional termstructure that incorporates the credit spreads.

Naturally, term structure models allowing for jumps are a field strongly related to credit risk. These models incorporate jumps in the dynamics of the term structure, and therefore the bond price processes also allow for discontinuties. In our framework a jump of the term strucuture/bond price is related to a default event. Shirakawa (1991) investigates a bond model where the forward rate curve follows a multidimensional Poisson-Gaussian process. In this setting, he finds necessary and sufficient conditions for completeness of the financial market and derives explicitly the price of a call option. With Björk, Kabanov and Runggaldier (1997) marked point processes entered the interest rate theory as sources of discontinuities. Marked point processes are a generalization of multivariate Poisson processes.

In the following we formulate a model for the stochastic behavior of corporate bond prices. In this context, corporate bonds are bonds issued by public liability companies or other legal entities. A public liability company is thereby a company whose shares are traded at the stock market. We focus on default risk and present a setup for explaining the yield spread between corporate bonds and government bonds by conditioning on a set of appropriate state variables where we use the Cox property. With government bonds we associate bonds that are free of default risk, or equivalently in this section free of credit risk.

As what the modeling of the default risk concerns, our model belongs to the intensity based approach. The behavior of the default events is modeled by a Cox process, i.e. a Poisson process with stochastic intensity. It is well-known that in such a framework, being under the equivalent martingale measure, the (spot) credit spread s of a corporate bond is the product of the stochastic intensity of a Cox process and the loss rate l of the company. Further, every defaultable bond is a contingent claim and its price process can be expressed by a conditional expectation under the equivalent martingale measure. Especially, the price of defaultable zero bond $v(\cdot, T)$ with maturity T can be expressed under an equivalent martingale measure by application of the Cox process property

$$v(0,T) = \mathbb{E}_Q \left\{ \exp \left(- \int_0^T r(t) + s(t)\, dt \right) \right\} ,$$

where r is the short rate, $s = l\, \lambda_Q$ is the (short) spread and λ_Q is the intensity of the Cox process under the equivalent martingale measure Q. For an overview of these results see for instance Lando (1997). Further, more general results on Cox processes are discussed by Rolski, Schmidli, Schmidt and Teugels (1999), and Grandell (1997) who gives a detailed characterization of the Cox process and studies its properties and discusses some special cases that are important in insurance mathematics. The martingale aspects of the Cox process is emphasized by Brémaud (1981).

The goal of this article is to describe totally the credit spread s of a corporate bond by "explaining factors". Therefore we define an appropriate environment. Denote $(\mathcal{G}_t)$ the sub-market filtration explaining the credit spread; the (spot) credit spread s is a $(\mathcal{G}_t)$-predictable process. The credit spread explaining sub-market filtration $(\mathcal{G}_t)$ is generated by the price processes of a riskless money market account, of a riskless zero bond (both free of default risk) and of the company's stock price process. Here, we focus on modeling the (spot) credit spread s. Thus, a mathematical convenient choice for introducing credit risk in the market is a defaultable money market account C (see also Schönbucher (2000)). We assume that C exhibits a (negative) jump whenever a default occurs.

5.2 The Market Model

In this section we present the market model and introduce some required assumptions. They are assumed to hold from now on if not stated otherwise.

The modeling takes place in an intensity based framework; i.e. default is triggered by a point process with an intensity λ. The default intensity λ can be seen as a function of certain describing variables. Examples of such variables in our setup are the short rate process $r = \{r(s) : 0 \le s \le T\}$ and the stock price process $S = \{S(s) : 0 \le s \le T\}$ of the company.

As mentioned in the introduction, we consider a market model which consists of a money market account and a zero bond (both free of default risk), a company's stock and a defaultable money market account issued by the same company. This market model is set in a filtered probability space $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{0\leq t\leq T}\right)$ large enough to support a two dimensional standard Brownian motion $W = (W^1, W^2)$, $W^i = \{W^i(t) : 0 \leq t \leq T\}$, $i = 1, 2$, and a non-explosive point process $N = \{N(t) : 0 \leq t \leq T\}$, where $T > 0$ is some finite time horizon.

In what follows we need the information structures

$$\begin{aligned}\mathcal{F}_t &\equiv \sigma\left(W^1(s), W^2(s), N(s) : 0 \leq s \leq t\right),\\ \mathcal{G}_t &\equiv \sigma\left(W^1(s), W^2(s) : 0 \leq s \leq t\right),\end{aligned}$$

and

$$\mathcal{C}_t \equiv \mathcal{F}_t \vee \mathcal{G}_T, \qquad \text{for } 0 \leq t \leq T.$$

In the first two cases, we always think of them as the augmentation of the natural filtration, and $(\mathcal{C}_t)$ is also understood as the augmentation of the given filtration. For the exact definition of *augmentation* see Karatzas and Shreve (1991), p. 89. Naturally, for some technical reason, we take the continuous version of the Brownian motion W and the right continuous version of the point process N. This ensures that the *usual hypotheses* hold. Note that these conditions are necessary for stochastic integration with respect to semimartingales, see Protter (1990), p. 3.

The filtration $(\mathcal{F}_t)$ is the *market filtration*, whereas $(\mathcal{G}_t)$ is the *credit spread explaining sub market filtration*; i.e. $\mathcal{G}_t \subseteq \mathcal{F}_t$ for all $0 \leq t \leq T$. The sigma field $\mathcal{G}_T$ is used to define the *Cox filtration* $(\mathcal{C}_t)_{0\leq t\leq T}$. We have $\mathcal{C}_t = \mathcal{F}_t \vee \mathcal{G}_T = \mathcal{F}_t^N \vee \mathcal{G}_T = \mathcal{F}_t^N \vee \mathcal{C}_0$, where $(\mathcal{F}_t^N)$ is the natural filtration of the point process N, moreover $\mathcal{G}_t \subseteq \mathcal{F}_t \subseteq \mathcal{C}_t$, for $0 \leq t \leq T$. Here, we refer to the Appendix for the definition (and existence) of a Cox process in our kind of setup. Moreover, let $\Lambda = \{\Lambda(t) : 0 \leq t \leq T\}$ denote the compensator of N with respect to $(\mathcal{F}_t)$. Thus Λ is a $(\mathcal{F}_t)$–predictable process with paths of finite variation and $N - \Lambda$ is a local $(\mathcal{F}_t)$–martingale. For the definition of a compensator see, e.g., Protter (1990), Ch. III, p. 97.

Next we describe our market model in more detail. The term structure is given by the money market account $B = \{B(t) : 0 \leq t \leq T\}$

$$B(t) \equiv \exp\left(\int_0^t r(u)\, du\right), \qquad \text{for } 0 \leq t \leq T, \tag{5.1}$$

and the zero bond $p(\cdot, T) = \{p(t, T) : 0 \leq t \leq T\}$ with maturity $T > 0$

$$p(t, T) \equiv \exp\left(-\int_t^T f(t, u)\, du\right), \qquad \text{for } 0 \leq t \leq T. \tag{5.2}$$

Both quantities are free of default risk. Analogously to Heath, Jarrow and Morton (1986), we use a one factor model given by

$$f(s, t) \;\equiv\; f(0, t) + \int_0^s \alpha(u, t)\, du + \int_0^s \sigma(u, t)\, dW^1(u), \tag{5.3}$$

and

$$r(t) \equiv f(t,t), \qquad \text{for } 0 \le s \le t \le T, \tag{5.4}$$

where we additionally define

$$A(t,T) \equiv \int_t^T \alpha(t,u)\,du\,, \tag{5.5}$$

and

$$D(t,T) \equiv \int_t^T \sigma(t,u)\,du\,, \qquad \text{for } 0 \le s \le t \le T. \tag{5.6}$$

We assume similar conditions as in Heath, Jarrow and Morton (1986).

Assumption 5.1 *(a) r is positive and*

$$\int_0^T r(u)\,du < \infty\,, \quad P\text{–a.s.}. \tag{5.7}$$

(b) $\alpha(\cdot,s)$ and $\sigma(\cdot,s)$ are progressively measurable with respect to the filtration $(\mathcal{G}_t)$ for $0 \le s \le T$, and σ is strictly positive.

(c) The objects $\alpha(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ allow to interchange the order of integration for P-a.a. $\omega \in \Omega$.

The stock price process $S = \{S(t) : 0 \le t \le T\}$ is defined by its initial value $S(0)$ at time zero and the stochastic differential equation

$$dS(t) = \mu(t)S(t)\,dt + \nu(t)S(t)\,dW^2(t)\,, \qquad \text{for } 0 \le t \le T. \tag{5.8}$$

By Itô's formula we find

$$S(t) = S(0)\,\exp\left(\int_0^t \mu(u)\,du - \frac{1}{2}\int_0^t \nu(u)^2\,du + \int_0^t \nu(u)\,dW^2(u)\right), \qquad \text{for } 0 \le t \le T. \tag{5.9}$$

Finally, we define the price process $C = \{C(t) : 0 \le t \le T\}$ of the defaultable money market account by

$$C(t) \equiv \Pi(t)\,\exp\left(\int_0^t (r(u) + s(u))\,du\right), \qquad \text{for } 0 \le t \le T, \tag{5.10}$$

where $s = \{s(t) : 0 \le t \le T\}$ is the (spot) spread process and $\Pi = \{\Pi(t) : 0 \le t \le T\}$ describes the loss fraction or negative return of the invested money after default events. The process Π is modeled by

$$\Pi(t) \equiv \prod_{0<u\le t} (1 - l(u)\,\Delta N(u)) = \prod_{n=1}^{N(t)} (1 - l(T_n))\,. \tag{5.11}$$

The jump times $(T_n)_{n\geq 1}$ of the point process N are associated with a default event of the company. The loss ratio process $l = \{l(t) : 0 \leq t \leq T\}$ takes values in the open interval $(0, 1)$. Therefore, at every default time T_n the defaultable money market account C bears a loss of $l(T_n)$ in fraction. Note that $l = 1$ is not possible in our setting. If we allowed $l = 1$ then the price process C could reach the absorbing state 0 which might cause some technical problems. In what follows we want C to be strictly positive.

In order to ensure the existence of S and C as semimartingales we have to presume some additional technical assumptions. Part (c) of the following assumption guarantees that N is a point process in the spirit of Brémaud (1981), whereas (d) leads one step further to Cox processes or so–called doubly stochastic processes (Brémaud, 1981, and Grandell, 1997).

Assumption 5.2 *(a) The processes μ and ν are progressively measurable with respect to $(\mathcal{G}_t)$. The process ν is strictly positive and*

$$\int_0^T |\mu(u)|\,du \quad + \quad \int_0^T \nu(u)^2\,du \quad < \quad \infty\,, \quad P\text{–}a.s.. \tag{5.12}$$

(b) The processes s and l are $(\mathcal{G}_t)$-predictable. The process s is strictly positive, the process l takes values in the open interval $(0, 1)$ and

$$\int_0^T s(u)\,du \quad < \quad \infty\,, \quad P\text{–}a.s.. \tag{5.13}$$

(c) Λ is absolutely continuous and has the representation

$$\Lambda(t) = \int_0^t \lambda(u)\,du\,, \qquad \text{for } 0 \leq t \leq T, \tag{5.14}$$

where $\lambda = \{\lambda(t) : 0 \leq t \leq T\}$ is a $(\mathcal{G}_t)$-predictable and strictly positive process satisfying

$$\int_0^T \lambda(u)\,du < \infty, \quad P\text{–}a.s. \tag{5.15}$$

(d) Moreover, we assume Λ is the compensator of N with respect to the Cox filtration $(\mathcal{C}_t)$; i.e. $N - \Lambda$ is a local martingale with respect to the filtration $(\mathcal{C}_t)$.

Remark. (1) Part (c) and (d) of Assumption 5.2 for the point process N lead naturally to doubly stochastic processes introduced by Cox (1955). Brémaud (1981) shows that N on the filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{C}_t))$ has conditional independent increments; that is $N(t) - N(s)$ is P-independent of $\mathcal{C}_s$ given $\mathcal{C}_0$. Since $\mathcal{G}_T = \mathcal{C}_0$ is the σ–algebra generated by W, we say N is driven by W. Moreover, we have for $k \in \mathbb{N}_0$ and $0 \leq s \leq t \leq T$

$$P\left(N(t) - N(s) = k \,|\, \mathcal{C}_s\right) = \exp\left(-\int_s^t \lambda(u)\,du\right) \frac{\left(\int_s^t \lambda(u)\,du\right)^k}{k!}\,. \tag{5.16}$$

Therefore N conditioned on $\mathcal{G}_T$ (in short hand notation $N|_{\mathcal{G}_T}$) can be interpreted as an inhomogeneous Poisson process.

The choice of N as a doubly stochastic process is very particular. Elliot, Jeanblanc and Yor (2000) discuss a more general approach. They consider a Brownian motion W and a single default event carried by a random time τ; i.e. $N(t) = \mathbf{1}_{\{\tau \leq t\}}$. The natural filtrations in their setup are $(\mathcal{G}_t)$ of W and $\left(\mathcal{F}_t^N\right)$ of N, whereas they denote $(\mathcal{F}_t) = \left(\mathcal{G}_t \vee \mathcal{F}_t^N\right)$ as the enlarged filtration. In their article the problem of preserving the martingale property of W on $(\mathcal{F}_t)$ is studied. They introduce the (H) hypothesis; that is, every square integrable $(\mathcal{G}_t)$-martingale is a square integrable $(\mathcal{F}_t)$-martingale. Whether the (H) hypothesis holds or not depends on the question whether W is a martingale on $(\mathcal{F}_t)$ or not. Here, we do not consider this problem since we assume W is a Brownian Motion on $(\mathcal{F}_t)$ and hence a $(\mathcal{F}_t)$–martingale. Naturally, W is still a Brownian motion on its natural filtration $(\mathcal{G}_t)$. In other words, the (H) hypothesis holds in our setting. This can be directly inferred from Section 5.6. In our case, we study the relation between $(\mathcal{F}_t)$ and $(\mathcal{C}_t)$ where we focus on preserving the Cox process property after a change of measure.

5.3 Representation Lemma and Girsanov's Theorem

In this section we represent the discounted price processes as Doléans Dade exponentials using classical stochastic integration theory (see e.g. Protter, 1990). Moreover, a version of Girsanov's theorem is presented which is adequate for our purposes. This result can be found in Björk, Kabanov and Runggaldier (1997). Last but not least, we study the problem of maintaining the Cox property under change of measure.

We start by rewriting the actual price processes as Doléans Dade exponentials.

Lemma 5.3 *Under the assumptions 5.1 and 5.2, we have for* $0 \leq t \leq T$

$$B(t) = \exp\left(\int_0^t r(u)\,du\right) = \mathcal{E}\left(R\right)(t)\,, \tag{5.17}$$

$$p(t,T) = p(0,T)\,\mathcal{E}\left(R_p\right)(t)\,, \tag{5.18}$$

$$S(t) = S(0)\,\mathcal{E}\left(R_S\right)(t)\,, \tag{5.19}$$

and

$$C(t) = \mathcal{E}\left(R_C\right)(t)\,, \tag{5.20}$$

where

$$R(t) \equiv \int_0^t r(u)\,du\,, \tag{5.21}$$

$$R_p(t) \equiv R(t) - \int_0^t A(u,T)\,du - \int_0^t D(u,T)\,dW^1(u) + \frac{1}{2}\int_0^t D(u,T)^2\,du\,, \tag{5.22}$$

$$R_S(t) \equiv \int_0^t \mu(u)\,du + \int_0^t \nu(u)\,dW^2(u)\,, \tag{5.23}$$

and

$$R_C(t) \equiv R(t) - \int_0^t l(u)\,dN(u) + \int_0^t s(u)\,du\,. \tag{5.24}$$

Proof. $B = \mathcal{E}(R)$ follows directly from (5.54). Equation (5.18) is a well-known result in interest theory, see e.g. Björk (1997). The representation of $p(\cdot, T)$ as a Doléans Dade exponential is a consequence of (5.53). $S(t) = S(0)\,\mathcal{E}(R_S)(t)$ follows from (5.47).

Therefore, it remains to prove (5.20). Since $l \cdot N$ is a process with $(l \cdot N)^c = 0$ P–a.s., we have by equation (5.55)

$$\Pi(t) = \prod_{0<u\le t} (1 - l(u)\,\Delta N(u)) = \mathcal{E}(-l \cdot N)(t), \qquad \text{for } 0 \le t \le T. \tag{5.25}$$

Plugging this in equation (5.10) and applying (5.53) and (5.57) yields for $0 \le t \le T$

$$\begin{aligned} C(t) &= \Pi(t) \exp\left(\int_0^t (r(u) + s(u))\, du\right) \\ &= \mathcal{E}(-l \cdot N)(t)\, \mathcal{E}\left(\int_0^\cdot (r(u) + s(u))\, du\right)(t) \\ &= \mathcal{E}(-l \cdot N)(t)\, \mathcal{E}\left(R + \int_0^\cdot s(u)\, du\right)(t) \\ &= \mathcal{E}\left(R - l \cdot N + \int_0^\cdot s(u)\, du\right)(t) \\ &= \mathcal{E}(R_C)(t), \end{aligned}$$

as we claimed before. □

Now, we define the discounted price processes.

Definition 5.4 *The processes* $Z_p = \{Z_p(t) : 0 \le t \le T\}$, $Z_S = \{Z_p(t) : 0 \le t \le T\}$ *and* $Z_C = \{Z_p(t) : 0 \le t \le T\}$ *are defined through*

$$Z_p(t) \equiv \frac{p(t,T)}{B(t)}, \; Z_S(t) \equiv \frac{S(t)}{B(t)} \;\; \textit{and} \;\; Z_C(t) \equiv \frac{C(t)}{B(t)}, \qquad \textit{for } 0 \le t \le T.$$

The next lemma states the representation of the discounted price processes in terms of Doléans Dade exponentials.

Lemma 5.5 (Representation Lemma) *With the notation in Definition 5.4 we have*

$$\begin{aligned} Z_p(t) &= p(0,T)\,\mathcal{E}(Y_p)(t), \\ Z_S(t) &= S(0)\,\mathcal{E}(Y_S)(t), \end{aligned}$$

and

$$Z_C(t) = \mathcal{E}(Y_C)(t), \qquad \textit{for } 0 \le t \le T,$$

where

$$\begin{aligned} Y_p(t) &\equiv \frac{1}{2}\int_0^t D(u,T)^2\,du - \int_0^t A(u,T)\,du - \int_0^t D(u,T)\,dW^1(u)\,, \\ Y_S(t) &\equiv \int_0^t \mu(u)\,du - \int_0^t r(u)\,du + \int_0^t \nu(u)\,dW^2(u)\,, \end{aligned}$$

and

$$Y_C(t) \equiv \int_0^t s(u)\,du - \int_0^t l(u)\,dN(u)\,, \qquad \text{for } 0 \le t \le T.$$

Proof. We use Lemma 5.3 and some properties of the Doléans Dade exponential, to prove the desired results. First observe that

$$\frac{1}{B} = \exp\left(-\int_0^{\cdot} r(u)\,du\right) = \mathcal{E}\left(-R\right). \tag{5.26}$$

Now, we apply equation (5.57) to the representation of the price processes in Lemma 5.3. This is possible since R is a process with P–a.s. continuous paths of finite variation. We derive that

$$\begin{aligned} Z_p &= p(0,T)\,\mathcal{E}\left(R_p\right)\,\mathcal{E}\left(-R\right) &&= p(0,T)\,\mathcal{E}\left(Y_p\right), \\ Z_S &= S(0)\,\mathcal{E}\left(R_S\right)\,\mathcal{E}\left(-R\right) &&= S(0)\,\mathcal{E}\left(Y_S\right)\,, \end{aligned}$$

and

$$Z_C = \mathcal{E}\left(R_C\right)\mathcal{E}\left(-R\right) = \mathcal{E}\left(R_C - R\right) = \mathcal{E}\left(Y_C\right), \tag{5.27}$$

which finishes the proof. □

The key to all needed equivalent measures lies in the following version of Girsanov's theorem (see Björk, Kabanov and Runggaldier (1997), Theorem 3.12).

Theorem 5.6 (Girsanov's theorem) *Suppose we have a point process* $N = \{N(t) : 0 \le t \le T\}$ *on a filtered probability space* $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{0\le t\le T}\right)$*, where* $\Lambda = \{\Lambda(t) : t \ge 0\}$ *is the compensator of* N*, i.e.* $M = \{M(t) : 0 \le t \le T\}$ *defined by the equation* $M \equiv N - \Lambda$ *is a local martingale. We assume,* N *has a predictable intensity* $\lambda = \{\lambda(t) : 0 \le t \le T\}$*; that is*

$$\Lambda(t) = \int_0^t \lambda(s)\,ds\,, \qquad \text{for } 0 \le t \le T.$$

Let $W = \left(W^1, ..., W^d\right)$ *be a Standard Brownian motion,* $W^k = \left\{W^k(t) : 0 \le t \le T\right\}$ *for* $k = 1, ..., d$*. We assume* $(\mathcal{F}_t)_{0\le t\le T}$ *is the* P*–extension of the natural filtration of* (W, N)*, furthermore* $\mathcal{F} = \mathcal{F}_T$*.*
Let $\psi = \left(\psi^1, ..., \psi^d\right)$ *be a* d*–dimensional predictable process,* $\psi^k = \{\psi^k(t) : 0 \le t \le T\}$

for $k = 1, ..., d$, and let $\phi = \{\phi(t) : 0 \leq t \leq T\}$ be a strictly positive predictable process satisfying

$$\int_0^T ||\psi(s)||^2 \, ds < \infty \quad \text{and} \quad \int_0^T |\phi(s) - 1| \, \lambda(s) \, ds < \infty \quad P\text{–a.s.} \tag{5.28}$$

Define the process $L = \{L(t) : 0 \leq t \leq T\}$ by

$$L \equiv \mathcal{E}\left(\sum_{k=1}^{d} \psi^k \cdot W^k + (\phi - 1) \cdot M\right).$$

Suppose $\mathbb{E}_P\{L(T)\} = 1$, then we can define a measure Q by

$$dQ = L(T) \, dP.$$

We then have, Q is a probability measure equivalent to P, $\widetilde{W} = \left(\widetilde{W}^1, ..., \widetilde{W}^d\right)$ defined by

$$\widetilde{W}^k(t) \equiv W^k(t) - \int_0^t \psi^k(s) \, ds, \qquad \text{for } 0 \leq t \leq T \text{ and } k = 1, ..., d,$$

is a Standard Brownian motion under Q and the process $\lambda_Q = \{\lambda_Q(t) : 0 \leq t \leq T\}$, where

$$\lambda_Q(t) \equiv \phi(t) \, \lambda(t), \qquad \text{for } 0 \leq t \leq T, \tag{5.29}$$

is the intensity of N under the measure Q.
Moreover, every probability measure Q equivalent to P has the structure above.

Remark. (2) The present version of Girsanov's theorem is based on the martingale representation theorem for our specific filtered probability space. In Björk, Kabanov and Runggaldier (1997), Remark 3.2, the suitable martingale representation theorem is given. The result was established on Corollary 4.31 in Jacod and Shiryaev (1987). In fact, Björk, Kabanov and Runggaldier investigate a more complex situation by introducing a marked point process.

In Assumption 5.2 (d) we introduced the Cox property by enlarging the filtration to $(\mathcal{C}_t)$ and assuming that $N - \Lambda$ remains a local martingale on the Cox filtration $(\mathcal{C}_t)$. Naturally, the question arises what kind of measure changes preserve the Cox property of N; i.e. Assumption 5.2 (d). A partial answer can be found with Girsanov-Meyer (Theorem 5.18).

Corollary 5.7 *With the notation of the preceding theorem, we assume N is a P-Cox process conditioned on $(\mathcal{G}_t)$; i.e. Assumption 5.2 (c) and (d) hold, where $(\mathcal{G}_t)$ is the completed natural filtration of the Brownian motion W.*
If ψ and ϕ are $(\mathcal{G}_t)$-predictable and $\mathcal{E}((\phi - 1) \cdot M))$ is in $\mathcal{H}^2(P)$ (see (5.60) for the definition), then N is a Q-Cox process conditioned on $(\mathcal{G}_t)$. The Q-intensity of N is given by $\lambda_Q = \lambda \, \phi$.

Proof. We need to show that the Q-analogue to Assumption 5.2 (c) and (d) hold. Theorem 5.6 implies that the $((\mathcal{F}_t), Q)$-intensity λ_Q of N equals $\lambda\,\phi$. The process λ_Q is $(\mathcal{G}_t)$-predictable since λ and ϕ are $(\mathcal{G}_t)$-predictable. Thus N admits the $((\mathcal{F}_t), Q)$-intensity λ_Q that is $(\mathcal{G}_t)$-predictable.

Next, we study the change of measure given by $dQ = L(T)\,dP$ on the Cox filtration $(\mathcal{C}_t) = (\mathcal{F}_t \vee \mathcal{G}_T)_{0\le t\le T}$. Define

$$Z(t) \equiv \mathbb{E}_P\left\{\frac{dQ}{dP}\middle|\,\mathcal{C}_t\right\}, \quad \text{for } 0 \le t \le T.$$

Assumption 5.2 (d) gives us the local martingale property of M on the Cox filtration $(\mathcal{C}_t)$ and equation (5.28) with Theorem 8 in Bremaud, Ch. II., gives us the local martingale property of $(\phi-1)\cdot M$. Hence, the Doléans Dade exponential of $(\phi-1)\cdot M$ is also a $((\mathcal{C}_t), P)$-local martingale due to Corollary 5.17. Moreover, $\mathcal{E}((\phi-1)\cdot M)$ is a square integrable $((\mathcal{C}_t), P)$-martingale, since it is in $\mathcal{H}^2(P)$. Furthermore,

$$L(T) = \mathcal{E}\left(\sum_{k=1}^{d}\psi^k\cdot W^k + (\phi-1)\cdot M\right)(T) = \mathcal{E}\left(\sum_{k=1}^{d}\psi^k\cdot W^k\right)(T)\,\mathcal{E}((\phi-1)\cdot M)(T)$$

because $[W, M] = 0$, since W has continuous paths and M has paths of finite variation a.s.. Further, the vector process ψ is $(\mathcal{G}_t)$-predictable and thus the Doléans Dade exponential of the "integrated" Brownian motion $\psi\cdot W$ is $(\mathcal{G}_t)$-adapted. With $\mathcal{G}_T \subset \mathcal{C}_0$, we find that for every $0 \le t \le T$

$$\begin{aligned} Z(t) &= \mathbb{E}_P\{L(T)\,|\,\mathcal{C}_t\} \\ &= \mathbb{E}_P\left\{\mathcal{E}\left(\sum_{k=1}^{d}\psi^k\cdot W^k\right)(T)\,\mathcal{E}((\phi-1)\cdot M)(T)\,\middle|\,\mathcal{C}_t\right\} \\ &= \mathcal{E}\left(\sum_{k=1}^{d}\psi^k\cdot W^k\right)(T)\,\mathbb{E}_P\{\mathcal{E}((\phi-1)\cdot M)(T)\,|\,\mathcal{C}_t\} \\ &= \mathcal{E}\left(\sum_{k=1}^{d}\psi^k\cdot W^k\right)(T)\,\mathcal{E}((\phi-1)\cdot M)(t). \end{aligned}$$

Hence

$$Z(t) = Z(0)\,\mathcal{E}((\phi-1)\cdot M)(t), \quad \text{for } 0 \le t \le T,$$

where $Z(0) = \mathcal{E}\left(\sum_{k=1}^{d}\psi^k\cdot W^k\right)(T)$ is a $\mathcal{C}_0$-measurable random variable.

We see $\Delta Z(t) = Z(t-)\,(\phi(t)-1)\Delta N(t)$. Therefore,

$$\frac{Z(t-)}{Z(t)} = \frac{Z(t-)}{Z(t-)+\Delta Z(t)} = \frac{Z(t-)}{Z(t-)+Z(t-)\,(\phi(t)-1)\Delta N(t)} = \frac{1}{1+\Delta N(t)(\phi(t)-1)} \tag{5.30}$$

The jump process N is a classical semimartingale. We choose the decomposition $N = M_P - A_P$, where $M_P \equiv \phi\cdot M$ is a $((\mathcal{C}_t), P)$-local martingale due to Assumption 5.2 (d) and $A_P \equiv N - \phi \cdot M$ is a P-FV By Theorem 5.18, we get that

$$\begin{aligned}
M_Q(t) &= M_P(t) - \int_0^t \frac{1}{Z(s)}\, d[Z, M_P](s) \\
&= M_P(t) - \int_0^t \frac{1}{Z(s)}\, d[Z(0) + Z_- \cdot ((\phi - 1) \cdot M), \phi \cdot M](s) \\
&= M_P(t) - \int_0^t \frac{Z(s-)}{Z(s)}\, d[(\phi - 1) \cdot M, \phi \cdot M](s) \\
&= M_P(t) - \int_0^t \frac{1}{1 + \Delta N(s)(\phi(s) - 1)}\, (\phi(s) - 1)\, \phi(s)\, d[M, M](s) \\
&= (\phi \cdot M)(t) - \int_0^t \frac{1}{1 + \Delta N(s)(\phi(s) - 1)}\, (\phi(s) - 1)\, \phi(s)\, dN(s) \\
&= (\phi \cdot N)(t) - (\phi \cdot \Lambda)(t) - \int_0^t \frac{(\phi(s) - 1)\, \phi(s)}{1 + \phi(s) - 1}\, dN(s) \\
&= (\phi \cdot N)(t) - \int_0^t \phi(s)\, \lambda(s)\, ds - \int_0^t (\phi(s) - 1)\, dN(s) \\
&= N(t) - \int_0^t \lambda_Q(s)\, ds
\end{aligned}$$

is a $((\mathcal{C}_t), Q)$-local martingale. Thus λ_Q is the $((\mathcal{C}_t), Q)$-intensity of N. It is unique since λ_Q is predictable. This conclusion is the Q-equivalent formulation of Assumption 5.2 (d). Therefore, N is a Q-Cox process conditioned on $(\mathcal{G}_t)$ with intensity λ_Q (by a combination of Theorem 4, Definition 7 and Theorem 9 in Brémaud (1981), Ch. II., pp. 25). □

5.4 Completeness and Contingent Claim Valuation

In the present section we show that the set of all equivalent martingale measures connected to our market model is a singleton.

For a brief repetition, a probability measure Q is an *equivalent martingale measure* in our market model if $P \sim Q$ and the discounted price processes Z_P, Z_S and Z_C are $(Q, (\mathcal{F}_t))$-martingales. It turns out that the uniqueness of an equivalent martingale measure implies that the market model is complete. In other words, for every contingent claim we can find a self-financing trading strategy such that the payoff at maturity can be replicated. For the exact definition of completeness in our setting see Corollary 5.12.

Proving existence and uniqueness of an equivalent martingale measure demands some additional technical assumptions which are summarized next.

Assumption 5.8 *(a) The process $L = \{L(t) : 0 \le t \le T\}$ defined by*

$$L \equiv \mathcal{E}\left(\psi^1 \cdot W^1 + \psi^2 \cdot W^2 + (\phi - 1) \cdot M\right),$$

is a square integrable P–martingale, where the processes $\psi^k = \left\{\psi^k(t) : 0 \le t \le T\right\}$, $k = 1, 2$, and $\phi = \{\phi(t) : 0 \le t \le T\}$ are given for every $0 \le t \le T$ by

$$\psi^1(t) \equiv \frac{\frac{1}{2}D(t,T)^2 - A(t,T)}{D(t,T)}, \tag{5.31}$$

$$\psi^2(t) \equiv \frac{r(t) - \mu(t)}{\nu(t)}, \tag{5.32}$$

and

$$\phi(t) \equiv \frac{s(t)}{l(t)\,\lambda(t)}, \tag{5.33}$$

and satisfy the regularity conditions (5.28) in Theorem 5.6.

(b) The process $\mathcal{E}\left((\phi - 1) \cdot M\right)$ is in $\mathcal{H}^2(P)$ (see (5.60) for the definition).

(c) The discounted price processes Z_p, Z_S and Z_C are $\mathcal{H}^2(P)$–semimartingales.

Existence of the processes ψ^1, ψ^2 and ϕ is ensured by Assumptions 5.1 and 5.2. The square integrable martingale properties for L and for the Doléans Dade exponential $\mathcal{E}\left((\phi - 1) \cdot M\right)$ are needed for some technical reason in the proof of the following main result.

Theorem 5.9 *In the defined market model the set of all equivalent martingale measures $\mathcal{Q}$ is a singleton; i.e. $\mathcal{Q} = \{Q\}$.*

Proof. Since all defined price processes are strictly positive we know by Corollary 5.17 that the discounted price processes are local martingales if and only if their stochastic exponents are local martingales. In what follows we have to find conditions to ensure the local martingale property for the stochastic exponents given in Lemma 5.5.

Let $\mathcal{P}$ denote the set of all probability measures equivalent to the 'original' measure P. We fix a measure $P^* \in \mathcal{P}$. For the processes ψ_{P^*} and ϕ_{P^*}, satisfying the regularity conditions (5.28) in Theorem 5.6, we have for every $0 \le t \le T$

$$Y_p(t) = \frac{1}{2}\int_0^t D(u,T)^2\,du - \int_0^t A(u,T)\,du - \int_0^t D(u,T)\,\psi^1_{P^*}(u)\,du - \int_0^t D(u,T)\,d\widetilde{W}^1(u)\,,$$

$$Y_S(t) = \int_0^t \mu(u)\,du - \int_0^t r(u)\,du$$

and

$$Y_C(t) = \int_0^t s(u)\,du - \int_0^t l(u)\,\lambda(u)\,\phi_{P^*}(u)\,du - \int_0^t l(u)\,dM_Q(u)\,,$$

where we just replaced the P-martingales by the drift transformed Q-local martingales $\widetilde{W}$ and M_Q with respect to P^* using Theorem 5.6. Here $M_Q = N - \int_0^\cdot \lambda_{P^*}(u)\,du$, and $\lambda_{P^*} = \lambda\,\phi_{P^*}$ is the intensity of N under P^*.

In each line, the last integral is a local martingale whereas all the Lebesgue integrals are continuous and hence predictable processes of finite variation. As mentioned before, our price processes are local martingales if and only if their stochastic exponents have this property. Further, by the unique decomposition of a special semimartingale, see e.g. Theorem 18, Ch. III, Protter (1990), the stochastic exponents are local martingales if and only if the Lebesgue integrals become zero. As conclusion, the equality system

$$\begin{aligned} 0 &= \frac{1}{2} D(\cdot,T)^2 - A(\cdot,T) - D(\cdot,T)\,\psi^1_{P^*}\,, \\ 0 &= \mu - r + \nu\,\psi^2_{P^*}\,, \end{aligned}$$

and

$$0 \;=\; s - l\,\lambda\,\phi_{P^*}\,, \qquad dP \otimes dt \text{ a.s.}$$

is a necessary and sufficient condition for the local martingale property of the discounted price processes.

Straightforward calculations yield that $\psi^1(t), \psi^2(t)$ and $\phi(t)$ defined in (5.31)-(5.33) is the unique solution $(\psi^1_{P^*}, \psi^2_{P^*}, \phi_{P^*})$ for the above equality system. Thus, due to Assumption 5.8, the measure dQ defined by

$$dQ = L_Q(T)dP\,,$$

with $L_Q(T) \equiv \mathcal{E}\left(\psi^1_Q \cdot W^1 + \psi^2_Q \cdot W^2 + (\phi_Q - 1)\cdot M\right)(T)$ is an equivalent probability measure with respect to P and the discounted price processes are local Q–martingales according to the considerations above.

Further, by Assumption 5.8, L_Q is P-square integrable and the discounted price processes are $\mathcal{H}^2(P)$–semimartingales. Therefore, Corollary 5.20 applies and the discounted price processes are Q–martingales.

We conclude that $Q \in \mathcal{Q}$, where $\mathcal{Q}$ denotes the set of equivalent martingale measures. Moreover, the derivation of the necessary and sufficient conditions yields uniqueness of Q in the sense that $P^* \in \mathcal{Q}$ implies $L_Q = L_{P^*}$ a.s., where L_{P^*} is the density of the change of measure from P to P^*. This completes the proof. □

The next result is an immediate consequence of Corollary 5.7 and Theorem 5.9.

Corollary 5.10 *N is a Q-Cox process conditioned on $(\mathcal{G}_t)$ with unique intensity λ_Q.*

Proof. Due to the definition of ψ^1, ψ^2 and ϕ, Assumption 5.1, 5.2 and 5.8, all conditions of Corollary 5.7 are satisfied and the statement follows. □

Corollary 5.11 *The underlying measure P is a martingale measure if and only if*

$$A(t,T) = \frac{1}{2}D(t,T)^2\,, \tag{5.34}$$
$$\mu(t) = r(t) \ \text{and} \tag{5.35}$$
$$s(t) = l(t)\,\lambda(t)\,, \qquad \text{for } 0 \le t \le T. \tag{5.36}$$

Proof. By Theorem 5.6, $P = Q$ if and only if $\psi^1 = \psi^2 = 0$ and $\phi = 1$. □

We are now ready to show the completeness of the market model. For the notation in the next corollary see e.g. Protter (1990), p. 134.

Corollary 5.12 *The market model is complete. More precisely, for every $\mathcal{F}_T$-measurable random variable X with $\mathbb{E}_Q\left\{(X/B(T))^2\right\} < \infty$ there exists a vector process $h = (h_P, h_S, h_C)$ with $h_P \in L(Z_P), h_S \in L(Z_S)$ and $h_C \in L(Z_C)$ such that the discounted value process of X defined by $V(t) \equiv \mathbb{E}_Q\{X/B(T)\,|\mathcal{F}_t\}$ for $0 \le t \le T$ satisfies*

$$V(t) = V(0) + \int_0^t h_P(s)dZ_P(s) + \int_0^t h_S(s)dZ_S(s) + \int_0^t h_C(s)dZ_C(s), \quad 0 \le t \le T\,, \tag{5.37}$$

and $V(T) = X/B(T)$.

Proof. Let X be an arbitrary random variable satisfying the assumptions of the corollary and V be the corresponding discounted value process. From the martingale representation theorem (see Remark 3.2 in Björk, Kabanov and Runggaldier, 1997) we know that V can be written as a stochastic integral with respect to $\widetilde{W}^1, \widetilde{W}^2$ and M_Q, i.e. for every $0 \le t \le T$

$$V(t) = V(0) + \int_0^t \psi^1(s)\,d\widetilde{W}^1(s) + \int_0^t \psi^2(s)\,d\widetilde{W}^2(s) + \int_0^t \phi(s)\,dM_Q(s)\,, \tag{5.38}$$

where $V(0) = \mathbb{E}_Q\{X/B(T)\}$, $\mathbb{E}_Q\{\int_0^T \|\psi(s)\|^2\,ds\} < \infty$ and $\mathbb{E}_Q\{\int_0^T |\phi(s)|^2\,\lambda(s)\,ds\} < \infty$. Note that by definition V is a uniformly and square integrable Q-martingale.

Next, we try to replace in (5.38) the integrators $d\widetilde{W}^1, d\widetilde{W}^2, dM_Q$ by dZ_P, dZ_S, dZ_C in an adequate way in order to get (5.37). For this purpose recall that from Lemma 5.5 and Corollary 5.7 the discounted price processes Z_P, Z_S and Z_C are given for every $0 \le t \le T$ under the equivalent martingale measure Q by

$$Z_P(t) = p(0,T)\mathcal{E}(D(\cdot,T)\cdot\widetilde{W}^1)(t)\,,$$
$$Z_S(t) = S(0)\mathcal{E}(\nu\cdot\widetilde{W}^2)(t)\,,$$

and

$$Z_C(t) = \mathcal{E}(-l\cdot M_Q)(t)\,.$$

Thus in all three cases the discounted price processes have the form $Z = Z(0)\,\mathcal{E}(Y)$, where $Z(0)$ is P-a.s. constant and $Y = H\cdot U$, where $H \in L(U)$. The integrability condition holds since $D(\cdot, T)$ is continuous, $\int_0^T \nu(s)^2 ds < \infty$ a.s. and l is bounded. Further, by Proposition 5.16 we have that $Y = \left(\frac{1}{Z}\right)_- \cdot Z$ and $\left(\frac{1}{Z}\right)_- \in L(Z)$ since $\left(\frac{1}{Z}\right)_-$ is left continuous. Therefore, by Theorem 21, Protter (1990), p. 135, we conclude that

$$U = \frac{1}{H} \cdot (H \cdot U) = \frac{1}{H} \cdot Y = \frac{1}{H} \cdot \left(\left(\frac{1}{Z}\right)_- \cdot Z \right) = \left(\frac{1}{H} \left(\frac{1}{Z}\right)_- \right) \cdot Z \equiv K \cdot Z \,,$$

and $K \in L(Z)$.

Next note that the discounted value process V is the sum of stochastic integrals of the form $\xi \cdot U$ where $\xi \in L(U)$. Again, by Theorem 21, Protter (1990), p. 135, we have that

$$\xi \cdot U = \xi \cdot (K \cdot Z) = (\xi\, K) \cdot Z \equiv h \cdot Z$$

and $h = \xi\, K \in L(Z)$, since $K \in L(Z)$ and $\xi \in L(U) = L(K \cdot Z)$. Every $\mathcal{F}_T$-measurable random variable X with $\mathbb{E}\left\{(X/B(T))^2\right\} < \infty$ can be thus duplicated by a self-financing strategy and the market model is complete. $\square$

In what follows, we discuss the assumption that a bank account C of the company exists. Such an assumption might cause some problems in calibrating the model. Nevertheless, the corollary of the next proposition shows that a zero bond issued by the company – hence affected by default risk – can be seen as a contingent claim in our framework.

Proposition 5.13 *Let the contingent claim X be an $\mathcal{F}_T$–measurable and P–square integrable random variable and let $\pi_X = \{\pi_X(t) : 0 \le t \le T\}$ be the corresponding price process of X, then*

(a) $$\pi_X(t) \quad = \quad \mathbb{E}_Q \left\{ \exp\left(-\int_t^T r(u)\, du \right) X \,\middle|\, \mathcal{F}_t \right\}, \qquad \textit{for } 0 \le t \le T.$$

(b) Moreover, if X has the representation $X = \Pi(T)\, Y$, where Y is a $\mathcal{G}_T$–measurable and P–square integrable random variable, then

$$\pi_X(t) \quad = \quad \Pi(t)\, \mathbb{E}_Q \left\{ \exp\left(-\int_t^T (r(u) + s(u))\, du \right) Y \,\middle|\, \mathcal{G}_t \right\}, \qquad \textit{for } 0 \le t \le T.$$

Proof. Part (a) is the standard martingale argument; where in (b) we mainly have to use the Cox property of N stated in Assumption 5.2. Corollary 5.10 implies that N is a Q-Cox process with intensity $\lambda_Q = \phi\, \lambda$. Define $M_Q \equiv N - \Lambda_Q$, where $\Lambda_Q \equiv \int_0^{\cdot} \lambda_Q(s)\, ds$. The process M_Q is a $((\mathcal{C}_t), Q)$-local martingale, where $\mathcal{C}_t = \mathcal{F}_t \vee \mathcal{G}_T$.

Since l is bounded on $(0, 1)$, the local martingale property is preserved for the process $l \cdot M_Q$. The Doléans Dade exponential of $-l \cdot M_Q$ is $Z \equiv \mathcal{E}(-l \cdot M_Q)$ and clearly a $((\mathcal{C}_t), Q)$-local martingale. Note that for every $0 \le t \le T$

$$Z(t) = \mathcal{E}(-l \cdot N + l \cdot \Lambda_Q) = \mathcal{E}(-l \cdot N)(t) \exp(\int_0^t l(u)\, \lambda_Q(u)\, du) = \Pi(t)\; \exp(\int_0^t s(u)\, du)$$

because $s = \phi\,\lambda\,l = \lambda_Q\,l$ and $[N, \Lambda_Q] = 0$ since Λ_Q is continuous and of finite variation. Without loss of generality, we may assume Z is a martingale. If this is not the case we find a sequence of stopping times (T_n) such that Z^{T_n} is a martingale for each n. The sequence (T_n) is a $(\mathcal{G}_t)$-stopping time and hence $\mathcal{C}_0$-measurable. This is possible, since $Z > 0$ and Z is bounded by the $\mathcal{G}_T$-measurable expression $\exp(\int_0^T s(u)\,du)$. Passing the limit $n \to \infty$ yields the same results by monotone convergence.

Using the fact that Z is a martingale and s is $(\mathcal{G}_t)$-adapted and thus $\mathcal{C}_0$-measurable, we derive that for every $0 \le t \le T$

$$\begin{aligned}
\mathbb{E}_Q\{\Pi(T)\,|\mathcal{C}_t\} &= \mathbb{E}_Q\left\{Z(T)\exp(-\int_0^T s(u)\,du)\,|\mathcal{C}_t\right\} \\
&= \mathbb{E}_Q\{Z(T)\,|\mathcal{C}_t\}\,\exp(-\int_0^T s(u)\,du) \\
&= Z(t)\,\exp(-\int_0^T s(u)\,du) \\
&= \Pi(t)\,\exp(-\int_t^T s(u)\,du)\,.
\end{aligned}$$

Therefore, for all $0 \le t \le T$

$$\begin{aligned}
\pi_X(t) &= \mathbb{E}_Q\left\{\exp\left(-\int_t^T r(u)\,du\right)\,Y\,\Pi(T)\,\middle|\,\mathcal{F}_t\right\} \\
&= \mathbb{E}_Q\left\{\mathbf{E}_Q\left\{\exp\left(-\int_t^T r(u)\,du\right)\,Y\,\Pi(T)\,\middle|\,\mathcal{C}_t\right\}\middle|\,\mathcal{F}_t\right\} \\
&= \mathbb{E}_Q\left\{\exp\left(-\int_t^T r(u)\,du\right)\,Y\,\mathbb{E}_Q\{\Pi(T)\,\,|\mathcal{C}_t\}\middle|\,\mathcal{F}_t\right\} \\
&= \mathbb{E}_Q\left\{\exp\left(-\int_t^T r(u)\,du\right)\,Y\,\Pi(t)\,\exp\left(-\int_t^T s(u)\,du\right)\middle|\,\mathcal{F}_t\right\} \\
&= \Pi(t)\,\mathbb{E}_Q\left\{\exp\left(-\int_t^T (r(u)+s(u))\,du\right)\,Y\,\middle|\,\mathcal{F}_t\right\}\,.
\end{aligned}$$

Finally, we need to show that we can replace $\mathcal{F}_t$ by $\mathcal{G}_t$ in the last expression. We define

$$U \equiv \exp\left(-\int_0^T (r(u)+s(u))\,du\right)\,Y\,.$$

The random variable U is $\mathcal{G}_T$-measurable, in $L^1(Q)$ and

$$\pi_X(t) = \Pi(t)\,\exp\left(\int_0^t (r(u)+s(u))\,du\right)\,\mathbb{E}_Q\{U\,|\,\mathcal{F}_t\}\,.$$

Note that the processes $\mathbb{E}_Q\{U\,|\,\mathcal{F}_t\}$ and $\mathbb{E}_Q\{U\,|\,\mathcal{G}_t\}$ are uniformly integrable martingales closed by the same random variable U and $\mathbb{E}_Q\{U\,|\,\mathcal{F}_T\} = U = \mathbb{E}_Q\{U\,|\,\mathcal{G}_T\}$. Thus a sufficient condition for the equality of these conditional expectation for all t is to show that $\mathbb{E}_Q\{U\,|\,\mathcal{G}_t\}$ is a $(\mathcal{F}_t)$-martingale. However, due to the martingale representation theorem for standard Brownian motions (see e.g. Theorem 42 in Protter 1990, p.155) we can represent $\mathbb{E}_Q\{U \mid \mathcal{G}_t\}$ as a stochastic integral with respect to the Q-Brownian motion $\widetilde{W}$. Note, that (W^1, W^2) and $(\widetilde{W}^1, \widetilde{W}^2)$ generate both the filtration $(\mathcal{G}_t)$, since $\widetilde{W}^k = W - \int_0^\cdot \psi^k(t)\,dt$ and ψ^k is adapted to the internal history of W, $(\mathcal{G}_t)$. Theorem 5.6 states that $\widetilde{W}$ is a Brownian motion on $(\mathcal{F}_t)$, hence $\mathbb{E}_Q\{U \mid \mathcal{G}_t\}$ is a $((\mathcal{F}_t, Q)$-martingale. □

Corollary 5.14 *Let $v(\cdot, T) = \{v(t,T) : 0 \leq t \leq T\}$ be the price process of the contingent claim $X \equiv \Pi(T)$ such that $v(\cdot, T)$ is a zero bond with default risk and maturity T. Then*

$$v(t,T) \quad = \quad \Pi(t)\,\mathbb{E}_Q\left\{\exp\left(-\int_t^T (r(u)+s(u))\,du\right)\middle|\,\mathcal{G}_t\right\}, \qquad \text{for } 0 \leq t \leq T.$$

Proof. Directly Proposition 5.13. □

Contingent claim valuation leads in such a setting directly to the well-known problem of pricing credit derivatives. Our setup belongs to the intensity based approaches, and the credit derivative pricing has been widely studied in such frameworks, see, e.g., Schönbucher (1998) and Lando (1997). In what follows we mainly focus on convertible bonds. For an introduction to convertible bonds see for instance Davis and Lischka (1999).

5.4.1 Convertible Bond Valuation

Convertible bonds are a combination of simple securities (bonds) and derivative securities. They are bonds which at the option of the holder can be converted into a specified number of common stock shares. They are referred to as hybrid securities since they contain both fixed income and equity components.

A convertible bond can be seen as the equivalent to the embedded corporate (default) bond plus an American option on the underlying stock with a changing strike price equal to the price of the embedded bond.

We assume that the stock pays no dividends. This is usually not restrictive since convertible bonds were originally developed for companies with poor credit. Such companies do not pay dividends. It is known that under this condition the pricing of a convertible bond simplifies to the pricing of a convertible bond which has European style, i.e. which can be only converted at the maturity $T_1 \leq T$. The payoff of a European convertible bond at time T_1 is in the above spirit given by

$$\begin{aligned} X_E &= v(T_1,T)1_{\{v(T_1,T)\geq c_0 S(T_1)\}} + c_0 S(T_1) 1_{\{v(T_1,T)< c_0 S(T_1)\}} \\ &= v(T_1,T) + (c_0 S(T_1) - v(T_1,T))_+\,, \end{aligned} \tag{5.39}$$

where $c_0 \geq 0$ denotes the number of shares specified at time $t = 0$ that can be converted at T_1. Note that if $c_0 = 0$ then we have just the contingent claim of a defaultable bond. Pricing a defaultable bond is therefore a special case of pricing a convertible bond.

The next lemma states once again the well-known property that the option of converting a convertible bond issued by a company not paying dividends before T_1 is worthless.

Lemma 5.15 *In our market model, the price process of a convertible bond with no dividend payments and maturity T_1 is given for every $0 \leq t \leq T_1$ by*

$$\begin{aligned} c(t,T_1) &= \mathbb{E}_Q\left\{ \exp\left(-\int_t^{T_1} r(u)du\right) X_E \,\middle|\, \mathcal{F}_t \right\} \\ &= v(t,T) + \mathbb{E}_Q\left\{ \exp\left(-\int_t^{T_1} r(u)du\right) \left(c_0 S(T_1) - v(T_1,T)\right)_+ \,\middle|\, \mathcal{F}_t \right\}. \quad (5.40) \end{aligned}$$

Proof. The result follows from standard type arguments since both assets are not subject to trading constraints and as well, include no additional cashflows, e.g., coupons and dividends. For completeness we briefly sketch the proof. Let $c_E(t,T_1), 0 \leq t \leq T_1$, be the price of a convertible bond without the option to convert before maturity (European style). Before proceeding, we claim that $c_E(t,T_1) \geq c_0 S(t)$ for all $0 \leq t \leq T_1$. If there exists $t \in [0,T_1]$ such that $c_E(t,T_1) < c_0 S(t)$ an arbitrage exists. To see this buy the convertible bond and sell short the stock for an initial cash flow of $c_0 S(t) - c_E(t,T_1) > 0$. Hold the position until the maturity of the bond. Convert the bond into stock and use this stock to cover the short position. The cash flow will be zero. Thus, the assumption that $c_E(t,T_1) < c_0 S(t)$ for every $t \in [0,T_1]$ gives us an initial positive cash flow with no risk of future loss. We conclude that $c_E(t,T_1) \geq c_0 S(t)$ for all $0 \leq t \leq T_1$.

Now we show that $c(t,T_1) \leq c_E(t,T_1)$ for all $0 \leq t \leq T_1$. If $c(t,T_1) > c_E(t,T_1)$ for some $0 \leq t \leq T_1$ an arbitrage exists. Buy the European convertible bond and sell the American bond for an initial cash flow of $c(t,T_1) - c_E(t,T_1) > 0$. If the counterparty converts before maturity, sell the European convertible bond and use the proceeds to purchase the stock required to cover the short position in the American convertible bond. Since we have shown that $c_E(t,T_1) \geq c_0 S(t)$ for all $0 \leq t \leq T_1$, the cash flow will be non-negative. If the counterparty holds until maturity, the two instruments are identical. Thus, the assumption that $c(t,T_1) \leq c_E(t,T_1)$ for all $0 \leq t \leq T_1$ gives us an initial positive cash flow with no risk of future loss. This is a contradiction to the assumption that the market is arbitrage free.

Finally, since the holder of the American convertible bond has all of the conversion opportunities as the holder of the European bond, it must be also that $c(t,T_1) \geq c_E(t,T_1)$ for all $0 \leq t \leq T_1$. By Proposition 5.13 and (5.39), the statement follows. □

5.5 A Martingale Model

We consider in this section an illustrative implementation of the continuous market model introduced in the previous sections under the equivalent martingale measure. Using the results which we developed so far, we compute numerically the fair price of a convertible bond.

Having evaluated theoretically the price process of a convertible bond in the last section, our objective is to compute and compare for illustrative purposes the fair price numerically for a particular martingale model which fits in our setting. We consider the market model directly under the equivalent martingale measure Q, i.e. all processes that we study are modeled under the equivalent martingale measure Q. In particular, all relations in Corollary 5.11 must hold. A very important task of the martingale modeling is the choice of the riskless spot and forward rate process in (5.4) and (5.3), respectively. We assume that the spot rate process for riskless debt is given by a Cox-Ingersoll-Ross model. In other words, $\{r(t) : 0 \leq t \leq T\}$ satisfies the stochastic differential equation

$$dr(t) = \alpha(\beta - r(t))dt + \sigma\sqrt{r(t)}dW^1(t) , \quad t \geq 0 , \tag{5.41}$$

where $r(0) = r_0 > 0$, $\alpha > 0, \sigma > 0$ and $\beta > \sigma^2/(2\alpha)$. In contrast to the Vasicek model, $\{r(t) : 0 \leq t \leq T\}$ given by (5.41) fulfills the condition $r(t) > 0$ a.s. for any $t \geq 0$ (see Assumption 5.1(a)). Moreover, the Cox-Ingersoll-Ross model still has nice computational properties such as the existence of a so called *affine term structure* (see e.g. Baxter and Rennie (1996) or Björk (1997)). By section 5.4 in Baxter and Rennie (1996) we conclude after some straightforward but tedious calculations that the Heath-Jarrow-Morton one factor model is completely specified by (5.41) and is given for every $0 \leq t \leq T$ by

$$\sigma(t,T) = \sigma^3\sqrt{r(t)}\,(\alpha + c(\alpha))\left(\frac{1}{2(\alpha + c(\alpha))} - \frac{1}{c(\alpha)}\right)\exp\left((\alpha + c(\alpha))(T-t)\right)Z(t,T)^{-2} \tag{5.42}$$

and

$$D(t,T) = \int_t^T \sigma(t,u)du = \sigma\sqrt{r(t)}\left(Z(t,T)^{-1} + \frac{c(\alpha)}{\sigma^2}\right) , \quad \text{respectively,} \tag{5.43}$$

where

$$Z(t,T) = -\frac{\sigma^2}{2(\alpha + \sigma^2 c(\alpha))} + \left(\frac{\sigma^2}{2(\alpha + \sigma^2 c(\alpha))} - \frac{1}{c(\alpha)}\right)\exp\left(-2(\alpha + c(\alpha))(T-t)\right)$$

and $c(\alpha) = -\alpha - \sqrt{\alpha^2 + 2\sigma^2}$. Recall once again that we are modeling under the equivalent martingale measure and hence the drift term $\alpha(t,T)$ and $A(t,T)$, for $0 \leq t \leq T$, can be easily established from (5.42) and (5.43), respectively, using Corollary 5.11. Further, by Proposition 3.5 of Björk (1997), the price process of the defaultable bond can be written for every $0 \leq t \leq T$ as

$$p(t,T) = \exp\left(-\alpha\beta\int_t^T (Z(u,T)^{-1} + c(\alpha))du - (Z(t,T)^{-1} + c(\alpha))r(t)\right) . \tag{5.44}$$

For the stock price process $S = \{S(t) : 0 \leq t \leq T\}$ in (5.8) we choose $\nu(t) = \nu$ whereas $\mu(t)$ is the riskless short rate $r(t)$.

It remains to specify the intensity of the default process $\{N(t) : 0 \leq t \leq T\}$. Again because of Corollary 5.11, modeling the intensity of the Cox process is in our framework equivalent with modeling the spread $s(t)$ and the loss rate $l(t)$, $t \geq 0$. Due to our market model, the quantities $s(t)$ and $l(t)$ can be described as functions of the short term process and the stock price process, i.e.

$$s(t) = f(\{r(u), S(u) : 0 \leq u \leq t\}) \quad \text{and} \quad l(t) = g(\{r(u), S(u) : 0 \leq u \leq t\})$$

for some measurable functions f and g. The properties of the functions f and g remain to be specified. Intuitively, it is clear that if the stock price process of the company is large and the negative price changes are small then default is very unlikely and the loss rate should also be small. Moreover, if the short rate is low the company can borrow money at a low rate of interest which is of course less risky and default is again not likely. For calibrating purposes, we assume that the functions f and g should not be too complicated.

The following linear approach takes our above considerations into account. We set for every $0 \leq t \leq T$

$$\begin{aligned} s(t) &= a_0 + a_1 r(t) + a_2 f_1(S(t)) + a_3 f_2(t, S)\,, \\ l(t) &= g(S(t))\,, \end{aligned} \tag{5.45}$$

where $S = \{S(u) : 0 \leq u \leq T\}, a_0, a_1, a_2, a_3 \geq 0$, $f_1 : \mathbb{R}^+ \to \mathbb{R}^+$ and $g : \mathbb{R}^+ \to (0,1)$ are non-increasing and $f_2 : [0,T] \times \mathcal{C}([0,T], \mathbb{R}_0^+) \to \mathbb{R}_0^+$. Examples for f_1 and g, respectively, are

$$\text{const}\,, \quad (1+x)^{-1}\,, \quad e^{-x}\,, \quad \text{or} \quad (1+\log(1+x))^{-1}\,, \quad x > 0\,,$$

whereas a natural choice for f_2 for every $0 \leq t \leq T$ and $x \in \mathcal{C}([0,T], \mathbb{R}^+)$ is for instance

$$1_{\{\inf\{\log x(u) - \log x(v) \,:\, u,v \in [0 \vee t-h, t]\} < z\}}\,, \quad \text{for } h > 0, z < 0 \ \text{ fixed}\,.$$

Figure 5.1 and 5.2 show numerical results for the price processes $\{c(t,T) : 0 \leq t \leq T_1\}$ and $\{v(t,T) : 0 \leq t \leq T_1\}$ (the case $c_0 = 0$) using the proposed setting for two examples. The implementation and testing of the suggested framework will be topic of a subsequent work and therefore the chosen parameters should be seen primarily as an illustration of our modeling.

The top two plots represent in both figures a simulated path of the stock price process $\{S(t) : 0 \leq t \leq T\}$ (left) and of the short rate process $\{r(t) : 0 \leq t \leq T\}$ (right). The dotted line in the right upper picture denotes the default adjusted short rate $\{r(t)+s(t) : 0 \leq t \leq T\}$. The spread $s(\cdot)$ was computed by formula (5.45) with $f_1(x) = (1+x)^{-1}$, $x > 0$, and $f_2(t,x) = 1_{\{\inf\{\log x(u) - \log x(v) \,:\, u,v \in [0 \vee t-0.4, t]\} < 0.18\}}$, $t \in [0,T]$, $x \in \mathcal{C}([0,T], \mathbb{R}_0^+)$. The chosen parameters $\nu, \alpha, \beta, a_0, a_1, a_2$ and a_3 are always summarized at the bottom of the plots.

The pictures in the second row in Figure 5.1 and 5.2 show the corresponding sample path of the Cox process (left) and of the stochastic intensity process (right). For simplicity, we set in both cases the recovery rate $l \equiv 0.5$.

The lower three plots in Figure 5.1 and 5.2 show the numerical results of our simulations. They indicate the corresponding price processes of a convertible bond with maturity $T_1 = 18$ and $T_1 = 19$, respectively, (solid line), of the defaultable zero-bond with maturity $T = 20$ and of the contingent claim $c_0 S(T_1)$ at time T_1 (dotted lines). Note that the price process of the contingent claim $c_0 S(T_1)$ is clearly given by $\{c_0 S(t) : 0 \leq t \leq T_1\}$ because of the martingale property of the stock price process under the equivalent martingale measure. The three plots are at each case generated by a different value of c_0 - the specified amount of common stock shares which can be obtained in the case of conversion. The upper two pictures present in both figures the price process of the convertible bond in a bullish market (convertible bonds behave more like stocks, left) and in a bearish market (convertible bonds behave more like bonds, right). The last plot shows the fair price for a particular c_0 in the region between the two economic extremes.

All diffusion processes were simulated by means of the Milstein scheme with stepsize $m = 0.1$ (strong Taylor approximation of convergence order 1) and we refer to Kloeden and Platen (1992) for details. The price processes for the convertible and defaultable bonds have been computed by Monte-Carlo simulation. Because of the large computation complexity we have chosen at each case only 100 simulations. However, the small number of simulations is justified by our results.

Figure 5.1 and 5.2 indicate that the sample paths of the price processes are reasonable smooth. Further, except in the bullish market, we can observe that the prices of convertible and defaultable bonds occur to have negative jumps whenever the corresponding Cox-process increases. This behavior is especially good visible in Figure 5.2. Note also that in Figure 5.1 all defaults happened because of the constant low level of the stock price process. The spread in Figure 5.1 is dramatically increasing in time. Considering the history of the stock price process the explanation might be that there is little hope that the stock price process will improve again in the future. Further defaults of the company are therefore likely.

Figure 5.2 shows a different scenario. Here, the first default appeared because of large negative changes of the (logarithmical) stock prices. The difference between the two consecutive prices at the time of the first default is -0.22. In contrast to the first case the spread is not increasing in time. The stock price process has developed differently. Although the stock price process is low at the end chances are still that it recover again.

In conclusion, we may say that our modeling approach yields a lot of feasible scenarios. The simulations produce realistic results and confirm that our modeling is reasonable.

5.6 Proofs, Definitions, and useful Results

Here, we give a short introduction to the stochastic analysis we need in this part of the dissertation. For details we refer to Protter (1990).

In the following, we work on a filtered probability space $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0}\right)$ satisfying the *usual hypotheses*. The processes $X = \{X(t) : t \geq 0\}$ and $Y = \{Y(t) : t \geq 0\}$ are semimartingales with $X(0) = Y(0) = 0$. The process $H = \{H(t) : t \geq 0\}$ is predictable.

First, we remind the reader of some notation. With the expression $X(t-)$ we associate $X(t-) = \lim_{s \nearrow t} X(s)$ for $t > 0$ and $X(0-) = 0$. The jump part $\Delta X = \{\Delta X(t) : t \geq 0\}$ of a semimartingale X is defined by $\Delta X(t) = X(t) - X(t-)$. If $\sum_{0<s\leq t} |\Delta X(s)| < \infty$ a.s., each $t \geq 0$, we can define the semimartingale $X^c = \{X^c(t) : t \geq 0\}$ which is the continuous part of the semimartingale X; i.e. $X^c(t) = X(t) - \sum_{0\leq s\leq t} \Delta X(s)$.

For the stochastic Itô integral we often use an abbreviation

$$H \cdot X \quad = \quad \int_0^{\cdot} H(s)\, dX(s). \tag{5.46}$$

Now, we recapture the definition of the Doléans Dade exponential. Let us consider the stochastic integral equation

$$Z(t) \quad = \quad 1 + \int_0^t Z(s-)\, dX(s)\,, \qquad \text{for } t \geq 0, \tag{5.47}$$

for a given semimartingale X. Equation (5.47) has the unique solution for $t \geq 0$

$$Z(t) = \exp\left(X(t) - \frac{1}{2}[X,X](t)\right) \prod_{0<s\leq t} (1 + \Delta X(s)) \exp\left(-\Delta X(s) + \frac{1}{2}\left(\Delta X(s)\right)^2\right), \tag{5.48}$$

alternatively,

$$Z(t) = \exp\left(X(t) - \frac{1}{2}[X,X]^c(t)\right) \prod_{0<s\leq t} (1 + \Delta X(s)) \exp\left(-\Delta X(s)\right)\,, \tag{5.49}$$

or, if $\sum_{0<s\leq t} |\Delta X(s)| < \infty$ a.s., for each $t \geq 0$,

$$Z(t) = \exp\left(X^c(t) - \frac{1}{2}[X,X]^c(t)\right) \prod_{0<s\leq t} (1 + \Delta X(s))\,. \tag{5.50}$$

The Doléans Dade exponential is also known as the *stochastic exponential of X*, written $\mathcal{E}(X)$.

Strictly positive semimartingales have the nice property that they can be represented as a Doléans Dade exponentials.

Proposition 5.16 *A strictly positive semimartingale Z with $Z(0) = 1$ allows the representation as Doléans Dade exponential $Z = \mathcal{E}(X)$, where X is unique; i.e.*

$$X(t) = \int_0^t \left(\frac{1}{Z}\right)(s-)\, dZ(s)\,, \qquad \textit{for } t \geq 0. \tag{5.51}$$

Corollary 5.17 *With the notation of the Proposition, Z is a local martingale iff X is a local martingale.*

The proof of the proposition is just the construction suggested by (5.51). The Corollary is a consequence of the equations (5.47) and (5.51), because Z is a stochastic integral with respect to X (with a càglàd integrand) and vice versa.

Equation (5.48) has some interesting special cases.

(a) If the semimartingale X has P–a.s. continuous paths, then

$$\mathcal{E}(X) = \exp\left(X - \frac{1}{2}[X,X]\right), \tag{5.52}$$

and

$$\exp(X) = \mathcal{E}\left(X + \frac{1}{2}[X,X]\right). \tag{5.53}$$

(b) If the semimartingale X has P–a.s. continuous paths of finite variation, then

$$\exp(X) = \mathcal{E}(X). \tag{5.54}$$

(c) If the semimartingale X is pure jump; i.e. $X^c = 0$, then

$$\mathcal{E}(X)(t) = \prod_{0<s\leq t}(1 + \Delta X(s)), \quad \text{for } t \geq 0. \tag{5.55}$$

Doléans Dade exponentials have also good properties with respect to multiplication.

(a) For two semimartingales X and Y with $X(0) = Y(0) = 0$ we have

$$\mathcal{E}(X)\,\mathcal{E}(Y) = \mathcal{E}(X + Y + [X,Y]). \tag{5.56}$$

(b) If X has P–a.s. continuous paths and either X or Y has paths of finite variation P–a.s. then

$$\mathcal{E}(X)\,\mathcal{E}(Y) = \mathcal{E}(X + Y). \tag{5.57}$$

A classical semimartingale X has decomposition $X = M+A$ where M is a local martingale and A is a FV process. This decomposition is in general not unique. The Girsanov-Meyer Theorem presents a possible decomposition of $X = L + C$ after a change of measure.

Theorem 5.18 (Girsanov-Meyer) *On a filtered probability space $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0}\right)$ let Q be an equivalent measure with respect to P and the process $Z = \{Z(t) : t \geq 0\}$ be defined by*

$$Z(t) \equiv \mathbb{E}_P\left\{\frac{dQ}{dP}\,\middle|\,\mathcal{F}_t\right\}, \qquad \textit{for } t \geq 0. \tag{5.58}$$

Let X be a classical semimartingale under P with decomposition $X = M + A$, where M is a local martingale and A is a FV process. Then X is also a classical semimartingale under Q and has a decomposition $X = L + C$, where

$$L(t) = M(t) - \int_0^t \frac{1}{Z(s)}\,d[Z,M](s), \qquad \textit{for } t \geq 0, \tag{5.59}$$

is a Q–local martingale and $C = X - L$ is a Q–FV process.

In the following, we consider a filtered probability space $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{0\le t\le T}\right)$ satisfying the *usual hypotheses* with a finite time horizon $T > 0$. Let X be a special semimartingale; i.e. X has a unique decomposition $X = M + A$, where M is a local martingale, A is a predictable process with paths of finite variation P–a.s. and $M(0) = A(0) = 0$. The $\mathcal{H}^2$ norm of X on the interval $[0, T]$ is defined by

$$||X||_{\mathcal{H}^2} = ||[M, M](T)^{1/2}||_{L^2} + || \int_0^T |dA(s)| \; ||_{L^2}. \tag{5.60}$$

The next theorem, Theorem 5 Chapter IV, Protter (1990), has a useful corollary.

Theorem 5.19 *Let X be a special $\mathcal{H}^2$–semimartingale, then*

$$\mathbb{E}\left\{\left(\sup_{0\le t\le T}|X(t)|\right)^2\right\} \quad < \quad 8\,||X||^2_{\mathcal{H}^2}.$$

Corollary 5.20 *Let X be a special $\mathcal{H}^2(P)$–semimartingale; i.e. X has a finite $\mathcal{H}^2$ norm under the measure P. Let L be a square integrable P–martingale defining the measure Q by $dQ = L(T)\,dP$. If X is a local Q–martingale, then X is a Q–martingale.*

Proof. By the inequality

$$\begin{aligned}
\mathbb{E}_Q\left\{\sup_{0\le t\le T}|X(t)|\right\} \quad &= \quad \mathbb{E}_P\left\{\sup_{0\le t\le T}|X(t)|\,L(T)\right\} \\
&\le \quad \frac{1}{2}\left(\mathbf{E}_P\left\{\left(\sup_{0\le t\le T}|X(t)|\right)^2\right\} + \mathbb{E}_P\left\{L(T)^2\right\}\right) \\
&\le \quad \frac{1}{2}\left(8\,||X||^2_{\mathcal{H}^2} + \mathbb{E}_P\left\{L(T)^2\right\}\right) \quad < \quad \infty
\end{aligned}$$

we have shown a sufficient condition for a local martingale to be a martingale, Theorem 47 Chapter I, Protter (1990). □

Finally, we recall the Definition of a Cox process. Therefore, we use a combination of Theorem 4, Definition 7 and Theorem 9 in Brémaud (1981).

Definition 5.21 *Let N be a non-explosive and adapted point process on a filtered probability space $\left(\Omega, \mathcal{C}, P, (\mathcal{C}_t)_{t\ge 0}\right)$ and λ a non-negative measurable process with*

$$\lambda(t) \textit{ is } \mathcal{C}_0\textit{-measurable and} \int_0^t \lambda(u)\,du < \infty\,, \textit{ for all } t \ge 0\,.$$

If $M \equiv N - \int_0^\cdot \lambda(u)\,du$ is a local martingale, then N is a Cox process with intensity λ.

Remark. (3) If the intensity process λ is predictable with respect to some filtration $(\mathcal{G}_t)$ and $\mathcal{G}_\infty \subset \mathcal{C}_0$, then we say N is a Cox process w.r.t. $(\mathcal{G}_t)$. Moreover, if λ is an appropriately measurable non-negative function f of some measurable process X; i.e. $\lambda(t) = f(t, X(t))$ and $\mathcal{G}_t = \mathcal{F}_t^X$, then we say N is a Cox process driven by X.

(4) Considering a Cox process w.r.t. some filtration $(\mathcal{G}_t)$ as in (3), we set $(\mathcal{F}_t) \equiv (\mathcal{F}_t^N \vee \mathcal{G}_t)$. By defining $(\mathcal{C}_t) \equiv (\mathcal{F}_t^N \vee \mathcal{G}_\infty)$ and the assumption that λ is the $(\mathcal{C}_t)$-intensity of N, we have N is a $(\mathcal{C}_t)$-Cox process. We say, N is a $(\mathcal{F}_t)$-Cox process conditioned on $(\mathcal{G}_t)$.

(5) The existence of a Cox process N conditioned on the natural filtration of a Brownian Motion W can easily be shown by a direct construction using Girsanov or even Girsanov-Meyer. We consider the case of a predictable intensity λ and a finite time horizon T. Suppose, we are given a standard Brownian Motion W on the filtered probability space $\big(\Omega^W, \mathcal{F}^W, P^W, (\mathcal{F}_t^W)\big)$ and a unit Poisson process on $\big(\Omega^N, \mathcal{F}^N, P^N, (\mathcal{F}_t^N)\big)$, where $(\mathcal{F}_t^W)$ and $(\mathcal{F}_t^N)$ are the natural filtrations. Taking the independent product of these two spaces, we get $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$, where $P \equiv P^W \otimes P^N$. Let $(\mathcal{C}_t) \equiv (\mathcal{F}_t^N \vee \mathcal{F}_T^W) = (\mathcal{F}_t \vee \mathcal{F}_T^W)$, then N is still a unit Poisson process on $(\mathcal{C}_t)$ since N and W are independent w.r.t. the common natural filtration $(\mathcal{F}_t)$.

For a $(\mathcal{F}_t^W)$-predictable positive process λ we define $L \equiv \mathcal{E}((\lambda - 1) \cdot M)$, where $M(t) \equiv N(t) - t$ is a $(\mathcal{C}_t)$-martingale. Suppose, $\mathbb{E}\{L(T)\} = 1$ then we can define the measure Q by $dQ \equiv L\, dP$ and $Q \sim P$. Moreover, we can show with Girsanov-Meyer that λ is the (unique) $((\mathcal{C}_t), Q)$-intensity of N (see Proof of Corollary 5.7). Further, $N - \int_0^\cdot \lambda(u)\, du$ is adapted to $(\mathcal{F}_t)$ and λ is $(\mathcal{F}_t)$-predictable, hence $N - \int_0^\cdot \lambda(u)\, du$ is a local $((\mathcal{F}_t), Q)$-martingale and λ is the $((\mathcal{F}_t), Q)$-intensity of N. A formal proof of this result can be carried out under some technical assumptions on λ.

5.7 Figures

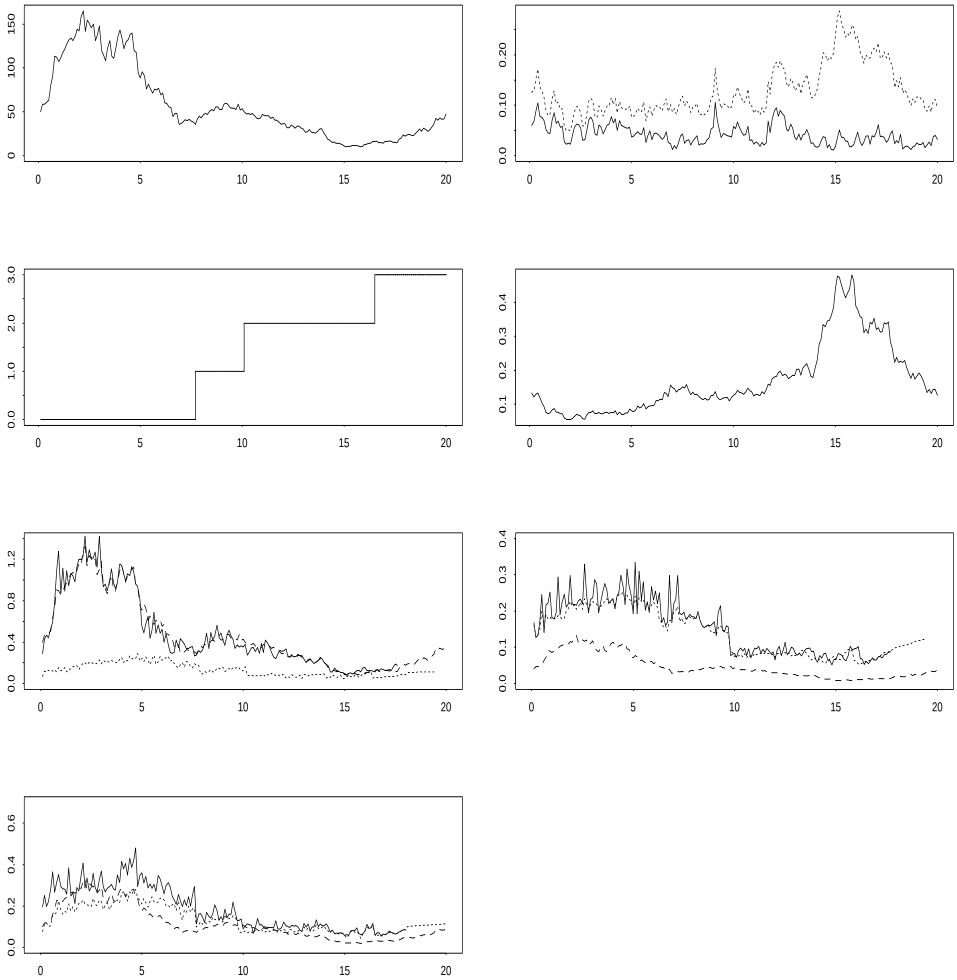

Figure 5.1: Scenarios for convertible bond valuation. Chosen parameters: $\nu = 0.3$, $\alpha = -2$, $\beta = 0.1$, $\sigma = 0.2$, $a_0 = 0.005$, $a_1 = 0.3$, $a_2 = 0.5$, $a_3 = 0$, $z = -0.18$, $l = 0.5$, $N = 100$, $m = 0.1$, $T = 20$, $T_1 = 18$ and $c_0 = 0.008$ (third row,left), 0.0008 (third row,right) and 0.002 (bottom,left), respectively.

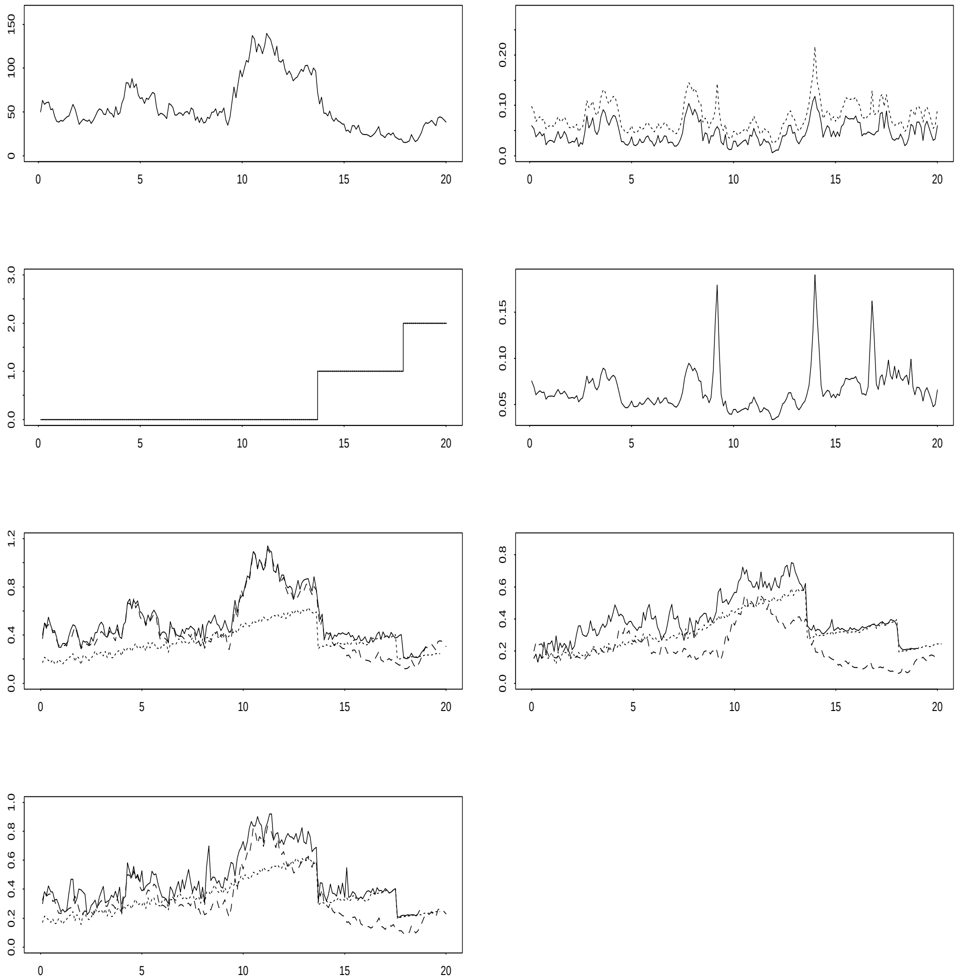

Figure 5.2: Scenarios for convertible bond valuation. Chosen parameters: $\nu = 0.3$, $\alpha = -2$, $\beta = 0.1$, $\sigma = 0.3$, $a_0 = 0.005$, $a_1 = 0.2$, $a_2 = 0.3$, $a_3 = 0.08$, $z = -0.18$, $l = 0.5$, $N = 100$, $m = 0.1$, $T = 20$, $T_1 = 19$ and $c_0 = 0.008$ (third row,left), 0.002 (third row,right) and 0.004 (bottom,left), respectively.

List of Tables

List of Figures

Bibliography

[1] Andrews, D.W.K. (1999a) Estimation when a Parameter is on a Boundary. In: *Econometrica* **67**, *No. 6*, pp. 1341-1383.

[2] Andrews, D.W.K. (1999b) Testing when a Parameter is on a Boundary of the maintained Hypothesis. Unpublished Manuscript, Cowles Foundation Discussion Paper No. 1229, *Yale University.*

[3] Back, K. (1992) Insider Trading in Continuous Time. In: *Review of Financial Studies* **2**, pp. 387-409.

[4] Bardhan, I., and X. Chao (1996) On martingale measures when asset returns have unpredictable jumps. In: *Stochastic Processes and their Applications* **63**, pp. 35-54

[5] Barndorff-Nielsen, O.E., T. Mikosch and S.I. Resnick, (Eds.) (2001) *Lévy Processes.* Birkhäuser, Boston.

[6] Barndorff-Nielsen, O.E. and N. Shephard (2001) Non-Gaussian Ornstein-Uhlenbeck Based Models and some of their Uses in Financial Economics. In: *Journal of Royal Statistical Society, Series B,* **63**, pp. 167–241.

[7] Barndorff-Nielsen, O.E., E. Nicolato, and N. Shephard (2002) Some recent developments in stochastic volatility modelling. In: *Quantitative Finance* **2**, pp. 11–23.

[8] Baxter, M., and A. Rennie (1996) *Financial Calculus.* Cambridge University Press, Cambridge.

[9] Bera, A. K., and M. L. Higgins (1993) ARCH Models: Properties, Estimation and Testing. In: *Journal of Economic Surveys* **7**, pp. 305-362.

[10] Bertoin, J. (1996) *Lévy Processes.* Cambridge University Press, Cambridge.

[11] Björk, T. (1997) Interest Rate Theory. In: *Financial Mathematics, Bressanone, 1996.* Runggaldier, W. (Ed.), Lecture Notes in Mathematics 1656, Springer, New York.

[12] Björk, T. (1998) *Arbitrage Theory in Continuous Time.* Oxford Press, Oxford.

[13] Björk, T., Y. Kabanov, and W. Runggaldier (1997) Bond Market Structure in the Presence of Marked Point Processes. In: *Mathematical Finance* **7**, pp. 211-239.

[14] Blanchet-Scalliet, C. and M. Jeanblanc (2002) Hazard rate for credit risk and hedging defaultable contingent claim, Working Paper, *Evry University, France.*

[15] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. In: *Journal of Econometrics* **31**, pp. 307-327.

[16] Bollerslev, T., R. Chou, and K. Kroner (1992) ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence. In: *Journal of Econometrics* **52**, pp. 5-59.

[17] Boswijk, H.P. (2001) Testing for a Unit Root with Near-Integrated Volatility. Tinbergen Institute Discussion Paper No. 01-077/4, *Amsterdam University, Netherlands.*

[18] Breiman, L. (1968) *Probability.* Addison-Wesley, Reading, Massachusetts.

[19] Brémaud, P. (1981) *Point Processes and Queues. Martingale Dynamics.* Springer, New York.

[20] Brooks, C., S. P. Burke, and G. Persand (2001) Benchmark and Accuracy of GARCH Model Estimation. In: *International Journal of Forecasting* **17**, pp. 45-56.

[21] Brown, P., and R. Da Silva Rosa (1997) Takeovers: Who Wins?, In: *Journal of the Australian Society of Security Analysts* **4**, pp. 2-5.

[22] Brown, P., and E. Yew (2002) How Do Investors Regard ESOs?, In: *Australian Accounting Review*, Forthcoming.

[23] Carr, P., and V. Linetsky (2000) The Valuation of Executive Stock Options in an Intensity-Based Framework. In: *European Finance Review* **4**, pp. 211-230.

[24] Chatfield C. (2001) *Time-Series Forecasting.* New York, Chapman&Hall.

[25] Christoffersen, P., J. Hahn, and A. Inoue (2001) Testing and Comparing Value-at-Risk Measures. In: *Journal of Empirical Finance* **8**, pp. 325-342.

[26] Cox, D. (1955) Some statistical methods connected with series of events. In: *Journal of the Royal Statistical Society* **B17**, pp. 129-164.

[27] Cuny, C.J., and P. Jorion (1995) Valuing Executive Stock Options with Endogenous Departure, In: *Journal of Accounting and Economics* **20**, pp. 129-164.

[28] Davies, R.B. (1977) Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative. In: *Biometrika* **64**, pp. 247-254.

[29] Davies, R.B. (1987) Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative. In: *Biometrika* **74**, pp. 33-43.

[30] Davis, M., and F. Lischka (1999) Convertible bonds with market risk and credit risk, Working Paper. Available at http://www.fam.tuwien.ac.at/ mark/.

[31] Davis, R., and W.T.M. Dunsmir (1996) Maximum Likelihood Estimation for MA(1) Processes with a Root near the Unit Circle. In: *Econometric Theory* **12**, pp. 1-29.

[32] Demos, A. and E. Sentana (1998) Testing for GARCH Effects: A one-sided Approach. In: *Journal of Econometrics* **86**, pp. 97-127.

[33] Duan, J.C. (1995) The GARCH Option Pricing Model. In: *Mathematical Finance* **5**, pp. 13-32.

[34] Duffie, D., and K.J. Singleton (1999) Modeling Term Structures of Defaultable Bonds. in: *Review of Financial Studies* **12**, Oxford University Press, Oxford, pp. 679-720.

[35] El Karoui, N., and L. Martellini (2001) Dynamic Asset Pricing Theory with Uncertain Time-Horizon, Working Paper, *Marshall School of Business, USA.*

[36] Elliot, R.J., M. Jeanblanc, and M. Yor (2000) On Models of Default Risk. In: *Mathematical Finance* **10**, pp.179-196.

[37] Embrechts, P., C. Klüppelberg, and T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[38] Engle, R.F. (1982) Autoregressive conditional Heteroscedasticity with Estimates of the United Kingdom Inflation. In: *Econometrica* **50**, pp. 987-1007.

[39] Far, H. (2000) Convergence of Likelihood Ratios for some Discontinuous Processes. Preprint, No. 600, *University of Paris 6, France.*

[40] Frey, R., and A. McNeil (2000) Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. In: *Journal of Empirical Finance* **7**, pp. 271-300.

[41] Grandell, J. (1997) *Mixed Poisson Processes.* Chapman & Hall, London.

[42] Hall, B.J., and K.J. Murphy (2000) Stock Options for Undiversified Executives, Working Paper, *Harvard Business School, USA.*

[43] Hamilton, J.D. (1994) *Time Series Analysis.* New Jersey, Priceton University Press.

[44] Harrison, J.M. and Pliska, S. (1981) Martingales and stochastic integrals in the theory of continuous trading. In: *Stochastic Processes and their Applications* **11**, pp. 215-260

[45] Heath, D., R. Jarrow, and A. Morton (1992) Bond pricing and the term structure of interest rates. In: *Econometrica* **60**, pp. 77-106

[46] Heyde, C.C. (1997) *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation.* Springer, New York.

[47] Huddart, S. (1994) Employee Stock Options. In: *Journal of Accounting and Economics* **18**, pp. 207-231.

[48] Hull, J.C. (1993) *Options, Futures, and other Derivatives, 2nd Ed.* Englewood Cliffs, Prentice Hall.

[49] Jacod, J. and P. Protter (2000) *Probability Essentials.* Springer, Heidelberg.

[50] Jacod, J., and A.N. Shiryaev (1987) *Limit Theorems for Stochastic Processes.* Springer, New York.

[51] Jarrow, R.A., D. Lando, and S.M. Turnbull (1997) A markov model for the term structure of credit risk spreads. In: *Review of Financial Studies* **10**, pp. 481-523

[52] Jeulin, T., and M. Yor (1979) Inégalités de Hardy, et faux-amis. In: *Séminaire de Probabilités XIII 1977/78 - Springer Lecture Notes in Mathematics* **721**. C. Dellacherie, P.A. Meyer, and M. Weil (Eds.), pp. 332-359.

[53] Johnson, S.A., and Y.S. Tian (2000) Indexed Executive Stock Options. In: *Journal of Financial Economics* **57**, pp. 35-64.

[54] Jorion, P. (2001) *Value at Risk: The new Benchmark for Managing Financial Risk, 2nd Ed.* New York, McGraw-Hill.

[55] Karatzas, I., and S.E. Shreve (1988) *Brownian Motion and Stochastic Calculus.* Springer, New York.

[56] Kloeden, P.E., and E. Platen (1992) *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin.

[57] Klüppelberg, C., R.A. Maller, M. Van De Vyver and D. Wee (2002) Testing for Reduction to Random Walk in Autoregressive Conditional Heteroscedasticity Models. In: *The Econometrics Journal*, Forthcoming.

[58] Kulatilaka, N., and A.J. Marcus (1994) Valuing Employee Stock Options. In: *Financial Analysts Journal* **50**, pp. 46-56.

[59] Kurtz, T., and P. Protter (1991) Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations. In: *Annals of Probability* **19**, pp. 1035-1070.

[60] Kusuoka, S. (1999) A Remark on Default Risk Models. In: *Advances in Mathematical Economics* **1**, pp. 69-82.

[61] Kyle, A.S. (1982) Continuous Auctions and Insider Trading. In: *Econometrica* **53**, pp. 1315-1335.

[62] Lambert, R.A., D.F. Larcker, and R.E. Verrecchia (1991) Portfolio Consideration in Valuing Executive Compensation. In: *Journal of Accounting Research* **29**, pp. 129-149.

[63] Lando, D. (1997) Modelling bonds and derivatives with default risk. In: *Mathematics and Derivative Securities*, Dempster, M., and S. Pliska, (Eds.), Cambridge University Press, Cambridge, pp. 369-393.

[64] Lando, D. (1998) On Cox Processes and Credit Risky Securities. In: *Review of Derivatives Research* **2**, pp. 99-120.

[65] Lehmann E.L. (1986) *Testing Statistical Hypothesis.* New York, Wiley.

[66] Liptser, R.S., and A.N. Shiryaev (1977) *Statistics of Random Processes I: General Theory.* Springer, Heidelberg.

[67] Liptser, R.S., and A.N. Shiryaev (1987) *Statistics of Random Processes II: Applications.* Springer, Heidelberg.

[68] Madan, D.B., P.P. Carr, and E.C. Chang (1998) The Variance Gamma Process and Option Pricing. In: *European Finance Review* **2**, pp. 79-105.

[69] Madan, D.B., and E. Seneta (1990) The Variance Gamma (VG) model for share market returns. In: *Journal of Business* **63**, pp. 511-524.

[70] Maller, R.A., R. Tan, and M. Van De Vyver (2002) How Might Companies Value ESOs?, In: *Australian Accounting Review*, Forthcoming.

[71] Nelson, D.B. (1990) ARCH Models as Diffusion Approximations. In: *Journal of Econometrics* **45**, pp. 7-38.

[72] Øksendal, B. (1995) *Stochastic Differential Equations: An Introduction with Applications.* Springer, New York.

[73] Powell, R.G. (1997) Modelling Takeover Likelihood, In: *Journal of Business Finance & Accounting* **24**, pp. 1009-1030.

[74] Protter, P. (1990) *Stochastic Integration and Differential Equations: A New Approach.* Springer, New York.

[75] Rolski, R., H. Schmidli, V. Schmidt, and J. Teugels (1999) *Stochastic Processes for Insurance and Finance.* Wiley, New York.

[76] Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge.

[77] Schönbucher, P.J. (1998) Term structure modelling of defaultable bonds. In: *Review of Derivatives Research* **2**, pp. 161–192.

[78] Schönbucher, P.J. (2000) Credit Risk Modelling and Credit Derivatives. PhD Dissertation, *University of Bonn, Germany.*

[79] Schwartz, E.S. (1997) The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging. In: *Journal of Finance* **52**, pp. 923-973.

[80] Schweizer, M. (1995) On the Minimal Martingale Measure and the Föllmer-Schweizer Decomposition. In: *Stochastic Analysis and Applications* **13**, pp. 573-599.

[81] Shirakawa, H. (1991) Interest rate option pricing with Poisson-Gaussian forward rate curve processes. In: *Mathematical Finance* **1**, pp. 77-94.

[82] Sørensen, M. (1997) Estimating Functions for Discretely Observed Diffusions: A Review. In: *Selceted Proceedings of the Symposium on Estimating Functions. IMS Lecture Notes - Monograph Series* **32**, Basawa, I.V., Godambe, V.P., and R.L. Taylor (Eds.), pp. 305-325.

[83] Stoddart, E.K. (2001) Options in Valuing Equity Compensation Benefits, In: *Australian Accounting Review* **11**, pp. 211-230.

[84] Vasicek, O.A. (1977) An Equilibrium Characterization of the Term Structure. In: *Journal of Financial Economics* **5**, pp. 177-188.

[85] Vu, H.T.V. and S. Zhou (1997) Generalizaton of Likelihood Ratio Tests under nonstandard Conditions. In: *The Annals of Statistics* **25**, *No. 2*, pp. 897-916.

[86] Woerner, J. (2001) Statistical Analysis for Discretely Observed Lévy Processes. PhD Thesis, *University of Freiburg, Germany.*