Financial and Econometric Models for Credit Risk Management

Zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
der Universität Fridericiana zu Karlsruhe

genehmigte

DISSERTATION

von

Dipl. Wi.-Ing. Bernhard Martin

Tag der mündlichen Prüfung: 12.05.2003

Referent: Prof. Dr. S.T. Rachev

Korreferent: Prof. Dr. M. Uhrig-Homburg

Karlsruhe (2003)

Preface

Over the years financial modeling has more and more turned away from the notion that financial returns follow a Gaussian process with independent increments. Instead properties have been detected in financial returns time series that negate the classical normal i.i.d. assumption. This has led to call for new methods.

This thesis work focuses on credit instruments and the behavior of their returns. It examines four phenomena observed in the time series of credit returns:

- Heavy-tailedness and peakedness.
- Time-varying volatility.
- Long memory (long-range dependence).
- Cointegration.

Additionally, it analyzes the *interdependence* among these phenomena.

On the contrary to known credit risk models based on the normal assumption, the model for price returns in this thesis will assume a *stable distribution*. As empirical studies show, the daily returns of a bond and the resulting credit risk obey a stable law, exhibiting leptokurtic, heavy-tailed, and skewed distributions. This leads to the application of *stable Value at Risk* measures as they provide a better fit compared to normal ones.

However, recent research has found that credit returns show certain patterns that give rise to the assumption of *autocorrelation*, *heteroscedasticity*, and *long-range dependence (LRD)* which enhances the need for a refinement of stable credit models applied so far.

ARCH-type processes are known to be capable of describing heteroscedasticity in financial time series. Although such heteroscedastic processes exhibit heavy-tailedness even under the Gaussian assumption, it is found later that the applications of ARCH-type processes and stable distributions are not mutually exclusive.

Long memory has been analyzed for financial time series already in the literature but mostly for high-frequency trading data of non-credit instruments. This thesis examines the presence of long memory in daily credit return data, and analyzes its occurrence in both signed returns and volatility process. Moreover, it is found out that this phenomenon can be attributed to a certain risk factor of the return-generating process.

Observing the market prices of different credit qualities (rating grades), a joint long-term behavior becomes visible: different credit qualities of equal maturity are found to be *cointegrated*. Thus, the concept of cointegration plays an important role when describing credit returns. Finally, all information should be embedded into a system of equations.

The *central objective* of this research is to build for the first time a model that integrates the description of all the above phenomena into one model that has the capability to describe the behavior of corporate credit returns for various credit ratings and maturities. Therefore the concept of *stable VAR (Vector Autoregression)* is applied for this model. Cointegrated VARs are estimated for a set of maturities, assuming the equations of the different credit grades to be simultaneous.

The model identifies several risk factors that have an impact on the credit returns. The long memory and the heteroscedasticity can finally be attributed to one certain factor only. In addition, it is found that the developed long-memory process captures the volatility clustering, and that a joint application of long-memory models and traditional ARCH-type models does not work. It is demonstrated that the stable cointegrated VAR with the incorporated long-memory model exhibits the best performance in forecasting Value at Risk among the compared models.

Due to these results it can be concluded that stable long-memory models for the description of credit returns will certainly play a greater role in the future.

As a further result of the long-memory based stable cointegrated VAR, the tail correlations in the model's innovations have disappeared. This allows the application of a tractable Gaussian copula with stable marginals to model the dependence between the different rating grades (copulas provide a much better description of dependence than traditional measures such as correlation, however, their application to high-dimensional random vectors used to be difficult under the stable assumption in the past).

It turns out that the consideration and modeling of the mentioned phenomena leads to an improvement of the forecast accuracy for credit returns compared to traditional methods. This enables more precise estimations of Value at Risk.

Aside from this, the developed stable cointegrated VAR model for credit returns is supposed to have a broad range of usage such as scenario analysis, stress testing, and portfolio optimization. The latter will be addressed in the final chapter which serves as an outlook covering possible fields of application.

Contents

Chapter 1

Introduction: The Context Of Credit Risk Management

This chapter explains the need for sophisticated credit risk models. It briefly reviews the current state of the art in credit risk management and discusses the view of the regulators. Furthermore, it explains the characteristics of the three general types of credit models: structural model, reduced-form model, and hybrid model. Finally, the chief properties of the well known CreditMetrics$^{(\text{TM})}$ approach are introduced and its deficiencies with regard to the distributional assumption are discussed.

1.1 Overview

Traditional methods of credit risk management, i.e. strict underwriting standards, limit enforcement, and close counterparty monitoring, focus on pure risk reduction but cannot meet requirements such as risk measurement and the assessment of portfolio efficiency. Aside from these issues, credit risk managers have looked for a way to integrate all risk determinants such as exposure and probabilities of rating changes, or default into a single overall measure. Being such a measure, Value at Risk (VaR) has gained popularity over the years. Instead of merely measuring each obligor's risk separately, risk managers also have to consider portfolio effects such as concentration risk, which accounts for the correlations among different obligors.

The traditional way to achieve protection from concentration risk was to set exposure limits for certain obligor groups (e.g. industries). However, this neglected consideration of the relationship between risk and return. The portfolio view helps to apply methods of risk diversification. These are the steps towards a risk-based view of the capital allocation process.

Under the impact of a steadily increasing market for credit derivatives which are difficult to manage, there is a strong need for a quantitative approach. The use of derivatives has grown during the past years. Their exposure is not visible in the balance sheet of a financial institution and their structure is often extremely complex. Thus, the exposure due to credit derivatives is another reason for the application of *sophisticated quantitative methods* in credit risk management. In fact, the increasing use of such derivatives was the reason why the BIS (Bank for International Settlements) set up new requirements for risk-based capital in 1993.

Among the reasons why credit risk measurement in general has become more important over the past years is the fact that the number of bankruptcy cases has rocketed. This increase is due to structural reasons as it is caused by tight global competition. With the soar in the new economy companies (high-tech sector), which have brought with them an increasing need for capital on the one hand but face much higher risks on the other, the volume of high-yield bond markets is expected to grow further. The difficulties of firms within the new economy together with the growing number of bankruptcies underlines the riskiness of this segment.

All these effects result in a lower average credit quality in the credit market. Owing to ever sharper competition in the credit business, the margins for the creditors are declining, including the high-yield markets.

With increasing volatility in the markets, the values of collateral are more difficult to predict and are therefore, less likely to be realized. And with property values growing weaker and less safe, lending has become even riskier.

Advances in computer technology (especially in database technology) have supported the building of databases with historical loan/bond data and the application of complex quantitative methods such as multivariate Monte Carlo simulations.

Another major incentive for the development of new risk-based credit models is the fact that the regulations of the BIS and central banks are not satisfactory tools for adequate, risk-based decision making on capital allocation. The current legal requirements say that all loans to private sector companies must have the same capital reserve of 8%, regardless of the obligor's credit quality. Furthermore, according to the regulations, the capital requirements are simply added over all loans, disregarding possible portfolio effects such as diversification or concentration.

Starting in 1997, the regulators permitted a large number of banks to use internal models when calculating capital requirements for their trading books. In this case, the internal models applied have to observe certain constraints and have to be verified by backtesting. Such models also consider concentration risk and permit a precise calculation of the VaR for each trading instrument.

Internal models not only provide a measurement of credit risk and are a tool for the pricing of loans, bonds, and derivative instruments. Furthermore, they help with analyzing questions of capital allocation in the context of RAROC (risk adjusted return on capital).

The 1988 Basle Capital Accord[1] which determines the risk-based capital adequacy requirements imposed on banks, has become exposed to exploitation by banks as they have reduced their risk-based capital requirements, but not the actual risk within their portfolios. This development led to a game-theory-like exercise with banks frequently making use of regulatory capital arbitrage.[2]

The table below sets out the current risk-based capital (RBC) requirements:[3]

Type	Properties	Reserve
Loan	Uncollateralized/guaranteed	8.0 %
Loan	Collateralized/guaranteed	8.0 %
Loan	OECD government	0.0 %
Loan	OECD bank/securities dealer	1.6 %
Loan	Other collaterals/guarantees	8.0 %

Table 1.1: Overview of current risk-based capital (RBC) requirements for a number of banking book instruments.(OECD = Organization for Economic Cooperation and Development)

Nevertheless, because regulatory capital requirements are determined by the type of credit instruments, and not by the actual riskiness of the obligor, chances for capital arbitrage arise.

Regulators are aware of this problem, and thus, the Basle Committee proposed a framework for a new Capital Accord in January 2001.[4] It is the objective of the new Accord that the capital requirements that lending banks have to provide for their issued loans depend on the riskiness of the transaction. However, the debates on the current proposal are ongoing and further reviews will take place. The new Accord is expected to become effective by 2006.

Generally, the modeling of instruments which are subject to credit risk turns out to be a more complex task than the modeling of instruments which are driven merely driven by market risk. A higher number of risk sources exist for credit products. Credit risk models face difficulties that do not apply to market risk models.[5] Moreover, credit risk models usually differ from market risk models in that they focus on a longer time horizon.

The creditworthiness of an issuer is determined by various risk drivers: not only the issuer's particular financial well-being, but also economic developments and industry trends have an effect.

[1]Basle-Committee (1988).

[2]Saunders (1999, p. 6).

[3]Federal-Reserve-System-Task-Force (1998).

[4]Basle-Committee (2001).

[5]Huang (2000).

The main elements of credit risks are as follows:[6]

- Default probability.
- Recovery rate.
- Credit migration.
- Default and credit risk correlation.
- Risk concentration and credit concentration.

Credit risk models usually focus on spotlight credit events such as default or a change in the credit quality of the obligor. This criterion is used to classify the models as either *default-mode (DM) models* or *marked-to-market (MTM) models*. The former values credit instruments considering default as the only possible credit event, while the latter also takes the probabilities of changes in the obligor's credit rating into account. An MTM model can capture the likelihood of a decrease in the value of a credit instrument before default occurs. The credit instrument is valued as if it were traded in a market.

The credit event "default" happens very rarely. The event "change in credit rating" is also seldom observed and its timeliness depends largely on the reaction time of the institution assigning the rating (e.g. rating agency). Thus, concentrating on the behavior of daily credit prices and spreads could be a more successful way of describing the credit risk of an issuer. However, such an approach is limited to those instruments which are valued regularly by the market. In this case, the processes driving the price of a credit instrument need to be analyzed. As mentioned above, the prices of credit instruments are influenced by a wider range of factors than the prices of market instruments.

For traded credit instruments, market risk factors play a key role aside from individual counterparty risk. Therefore, it is critical to incorporate them into credit risk models. Moreover, as the impact of market movements is also relevant for short time horizons, credit risk models will adapt such horizons.

For non-traded instruments which are (usually) held to maturity (e.g. bank loans), DM-type models will continue to play a role. However, secondary credit markets are gaining importance and prices for loans in secondary markets depend on their current credit quality, the likelihood of future changes in credit quality, and market factors. Hence, DM-type models fail in this respect.

During times of financial crises, the close relationship between credit products and products exposed to market risk becomes visible, with the individual risk of the issuer remaining unaffected.[7]

[6] Ong (1999, p. 63).

[7] Huang (2000).

This demonstrates why credit risk models have to deal with more than counterparty risk itself. In this case, prices of tradable credit securities, market risk factors, and measures such as liquidity may also play a part. For example, when the US government announced that it was buying back old treasury bonds and reducing new issues, this caused an increased spread for long-dated corporates and a resulting higher liquidity premium.

Thus, credit spreads[8] are not driven solely by the issuer's default probability and the level of exposure. Specifically, when credit risk is measured on a marked-to-market basis, the prices of credit instruments interrelate closely with market interest rates (and other aggregated market factors).

Historically, credit spreads and interest rates have often moved in different directions. For example, when the markets feel that massive defaults are in the offing and investors switch to safe treasury bonds, then credit spreads increase, whereas interest rates drop due to increased demand for the treasury bonds.[9]

Apart from the distinction which credit events are considered by the model (DM and MTM models), credit risk models can be classified by another criterion: is the credit risk of a firm taken as an exogeneous factor or is it modeled depending on certain properties of the firm such as asset value? The following section presents an overview of the types of credit models classified according to this criterion. In all, three general forms are distinguished. The properties of each form are described and the major differences briefly discussed.

1.2 Structural, Reduced-Form, And Hybrid Models

Traditional approaches of credit risk models pinpointed single entities, and effects that occur within portfolios, were not analyzed. Analysis has dealt with expert systems, rating models, and credit scoring. Later models (see also the approach of CreditMetrics(TM) in section 1.3) targeted dependence and diversification effects among the obligors.

Today, *three* types of models can be distinguished: (i) *structural models*, (ii) *reduced-form models*, and (iii) *hybrid models*.

Structural models. The name comes from the underlying assumption that a firm's credit risk revolves around both its asset value and its balance sheet figures,

[8]Here credit spread is referred to as the difference between the yield achieved for a treasury bond and the yield achieved for corporate bonds of equal maturity.

[9]Huang (2000).

i.e. the structure of the company.[10] The first model of this kind was the Merton Model.[11] Structural models require assumptions on the development of the firm's asset values. This means the model has to build a link between the firm's asset-liability structure and its stock price movements.

Both default probability and recovery rate are contingent on these measures. As an extension, recovery rates may also be modeled exogeneously.[12]

Reduced-form models. Reduced-form models ignore the company's asset-liability structure or asset value. These models focus exclusively on the firm's traded liabilities (bonds) and the default-free term structure. The observed credit spread is seen as the result of the firm's default probability and expected recovery rate, both of which are exogeneous variables.

A disadvantage of structural models is that they assume the company's liability structure to be constant over time. This means that even if the value of corporate assets changes, the debt structure would be assumed to remain constant.[13] Reduced-form models are not affected by these considerations. However, reduced-form models are criticized for using prices taken from corporate debt markets, which are known to be less liquid than equity markets.

Hybrid models. Hybrid models incorporate properties of both structural and reduced-form models in an attempt to reap the benefits of both. A prominent example for hybrid models is J.P. Morgan's CreditMetrics(TM), which is described in the next section.

1.3 The CreditMetrics Approach

CreditMetrics(TM) was developed by J.P. Morgan and is a simulation-based tool for assessing portfolio risk due to changes in debt value.[14] The debt value of an obligor is influenced by two factors: (i) possible future downgrades in credit quality or (ii) the default probability of the obligor. The software tool can be applied to a number of different credit instruments such as loans, bonds, and credit derivatives. The measure for the risk of a portfolio over a given time horizon is its Value at Risk (VaR).

CreditMetrics(TM) is a hybrid model that merges a rating transition model with the assumption that joint credit quality moves are determined by joint movements

[10] Jarrow and Deventer (1999).
[11] Merton (1974).
[12] Longstaff and Schwartz (1995).
[13] Jarrow and Deventer (1999).
[14] Gupton, Finger and Bhatia (1997).

of the firm's assets.[15] The valuation approach is marked-to-market (MTM), which considers credit migration. This means that the value of a credit instrument is not determined exclusively by its default probability for future periods but also by the probabilities of changing credit quality. The current credit quality is represented by the credit rating (e.g. Moody's Investors Service, Standard & Poor's) of the obligor company. In this model, the risk of a company is assumed to be driven by its asset value. The correlations among credit events of different companies are calculated as correlations among their asset returns.[16]

In case of a default event, the creditor typically only receives a certain fraction of a bond's or loan's face value. This fraction is called *recovery rate*. In CreditMetrics$^{(TM)}$, the recovery rate is defined as a random variable.

One of the chief assumptions of CreditMetrics$^{(TM)}$ is that asset returns are normally distributed. The model uses a Merton-type approach that links asset value or the volatility of the returns to discrete rating migration probabilities of individual borrowers. The idea is that changes in the firm's asset value result in different probabilities for changing credit quality. The following graph depicts this concept.[17]

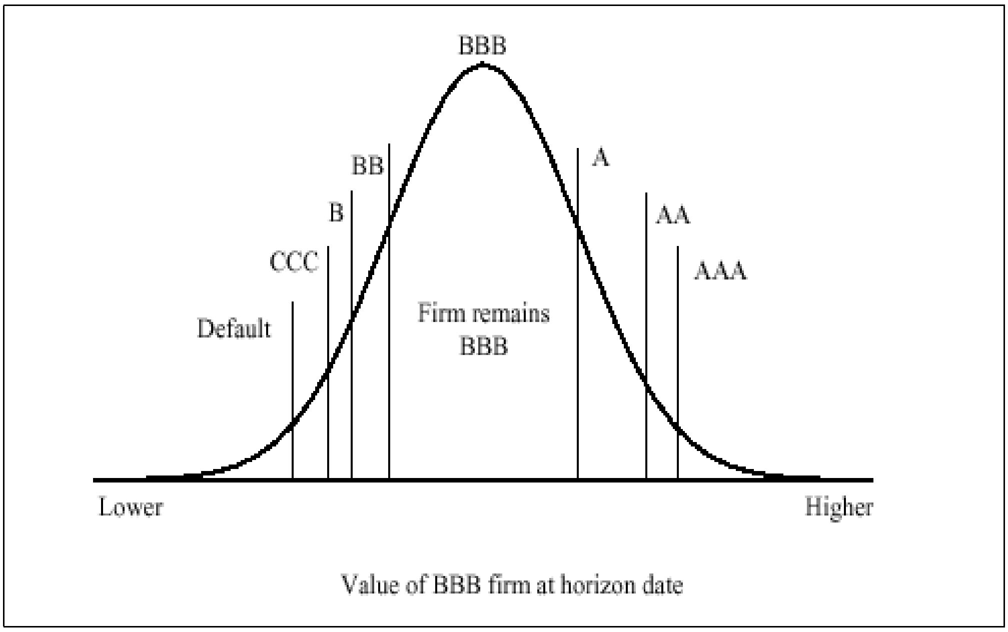

Figure 1.1: Mapping asset value and rating transition probabilities in CreditMetrics.

Figure 1.3 displays the (Gaussian) probability density of a company's future asset value. The thresholds on the value scale divide the space under the density

[15] Rachev, Schwartz and Khindanova (2001).
[16] Saunders (1999, p. 37).
[17] Figure taken from Gupton, Finger and Bhatia (1997).

function into different sectors. Each sector represents the probability of changing over into a certain credit rating category during the next period. These probabilities are derived from historical transition matrices. The method of calculating such thresholds for the asset value is explained in the CreditMetrics$^{(TM)}$ technical document and will not be gone into here.[18]

1.4 The Normal Assumption

The distributional assumption for financial returns and for the underlying risk factors has a considerable impact on the risk measures applied. For derivative instruments, the assumption of the type of distribution is essential in terms of pricing models.

In financial models, the distributions of asset returns were commonly assumed to follow a Gaussian law. Most of the financial theory is based on Bachelier's theory of speculation using the normal distribution for the asset returns.[19] B. Mandelbrot[20] and E. Fama[21] were the first to reject the Gaussian hypothesis and proposed the use of *stable distributions* for asset returns which provided a better fit on empirical samples.

The popularity of the normal distribution for modeling asset returns is based on its statistical simplicity and the Central Limit Theorem (CLT).[22] The assumption of independent identically distributed (i.i.d.) normal returns supports the Efficient Market Hypothesis.

The existence of a second moment has desirable properties: (i) the variance is a common measure used to express the individual riskiness of an asset and (ii) the dependence between the returns of two different assets are described by the covariance.

However, the Gaussian distribution is unable to capture *heavy-tailedness* of financial returns since its weight lies in the center. The focus of this thesis therefore centers on a broader class of distributions: *the stable distributions*.[23]

Stable distributions are able to capture not only the *heavy-tailedness and peakedness* of financial returns but also their *skewness*.

The following chapter explains the definition, the parameters, and the main properties of stable distributions. It also introduces stable vectors and demonstrates how to model dependence between the elements of a stable random vector.

---

[18] Gupton, Finger and Bhatia (1997).
[19] Bachelier (1900).
[20] See, for example, Mandelbrot (1997a), Mandelbrot (1997b) and Mandelbrot (1999).
[21] See, for example, Fama (1965a) and Fama (1965b).
[22] Feller (1971).
[23] The Gaussian distribution is a special case of the class of stable distributions.

Chapter 2

The Stable Distribution

This chapter first presents the definition, the parameters, and the properties of *stable distributions*. It goes on to define *stable random vectors* and the measure of *dependence* between stable random variables. Then it introduces *sub-Gaussian vectors* which are a method to represent the dependence between stable random vectors by applying Gaussian dependence measures. This method is of practical relevance. Moreover, measures that reflect the risk of stable asset returns are introduced: *variation* and *Value at Risk*. Finally, there is a brief overview of common financial performance measures defined under the stable assumption.

2.1 Definition And Parameters

Definition. There are several definitions of the stable distribution. The most common definition is as follows: a random variable X is said to have a stable distribution if, for any positive numbers A and B, there is a positive number C and a real number D, such that

$$AX_1 + BX_2 \stackrel{\mathrm{d}}{=} CX + D \;, \tag{2.1}$$

where X_1 and X_2 are independent copies of X, and where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. X is then called a stable random variable. If the equation holds with $D = 0$, then X is said to be *strictly stable*.[1]

The domain of attraction. Another definition for stable distributions that uses the concept of convergence in distribution for the sum of independent identically distributed random variables with infinite variance is the *domain of attraction*.

[1]For all definitions in this section see Samorodnitsky and Taqqu (1994, pp. 2).

A random variable X is said to have a stable distribution if it has a *domain of attraction*, i.e. if there is a sequence of i.i.d random variables $Y_1, Y_2, ..., Y_n$, and sequences of positive numbers $\{d_n\}$ and real numbers $\{a_n\}$ such that

$$\frac{Y_1 + Y_2 + ... + Y_n}{d_n} + a_n \overset{\mathrm{d}}{\Longrightarrow} X. \tag{2.2}$$

The notation $\overset{\mathrm{d}}{\Longrightarrow}$ means convergence in distribution.

The parameters of a stable distribution. Most stable distributions do not have a closed expression for their density function. However, all univariate stable distributions are defined by four parameters:

- α: Stability index $\in$ (0,2];
- β: Skewness parameter $\in$ [-1,1];
- σ: Scale parameter $\in R_+$;
- μ: Shift parameter $\in R$;

The Gaussian distribution commonly used in financial models is one member of the family of stable distributions. Gaussian distributions are characterized by $\alpha = 2$, while non-Gaussian stable distributions have an $\alpha \in (0, 2)$. In the case of $\alpha < 2$ the distributions show heavy tails and peakedness and do not have a finite variance. The tails decay like a power function, which also indicates that a stable random variable exhibits greater variability than a normal random variable.

Characteristic function. A random variable is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $\sigma \geq 0$,$-1 \leq \beta \leq 1$, and μ real such that its characteristic function has the following form:

$$Eexp\{i\theta X\} = exp\left\{-\sigma^\alpha|\theta|^\alpha\left(1 - i\beta(\operatorname{sign}\theta)\tan\frac{\pi\alpha}{2}\right) + i\mu\theta\right\} \text{ , if } \alpha \neq 1,$$

$$Eexp\{i\theta X\} = exp\left\{-\sigma|\theta|\left(1 + i\beta\frac{2}{\pi}(\operatorname{sign}\theta)\log|\theta|\right) + i\mu\theta\right\} \text{ , if } \alpha = 1; \tag{2.3}$$

where α is the index of stability, β is the skewness parameter, σ is the scale parameter, and μ is the shift parameter; sign(θ) is the sign-function with $sign(\theta) = +1$, for $\theta > 0$, $sign(\theta) = 0$, for $\theta = 0$, and $sign(\theta) = -1$, for $\theta < 0$. For a stable random variable X with a characteristic function (2.3) or (2.3), this is expressed as $X \sim S_\alpha(\sigma, \beta, \mu)$. If $\beta = 0$, then X is a *symmetric stable random variable* and denoted by $S\alpha S$.

2.2 Properties Of Stable Random Variables

Sum of two stable variables. Let X_1 and X_2 be two independent random variables with $X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i)$, $i = 1, 2$. Then, the parameters of the resulting distribution for $X_1 + X_2 \sim S_\alpha(\sigma, \beta, \mu)$ are given by

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha} \ , \tag{2.4}$$

$$\beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} \ , \tag{2.5}$$

$$\mu = \mu_1 + \mu_2 \ . \tag{2.6}$$

Extending this property to the case of more than two independent α-stable random variables X_i, $i = 1...n$, $n \geq 3$, the distribution $X = \sum_{i=1}^n X_i$ has the following parameters:

$$\sigma = \left(\sum_{i=1}^n \sigma_i^\alpha \right)^{1/\alpha} \ , \tag{2.7}$$

$$\beta = \frac{\sum_{i=1}^n \beta_i \sigma_i^\alpha}{\sum_{i=1}^n \sigma_i^\alpha} \ , \tag{2.8}$$

$$\mu = \sum_{i=1}^n \mu_i \ . \tag{2.9}$$

Multiplication with a constant. Another useful property of α-stable random variables is the scaling transformation. Let $X \sim S_\alpha(\sigma, \beta, \mu)$, and let a be a non-zero real constant. Then,

$$aX \sim S_\alpha(|a|\, \sigma, sign(a)\beta, a\mu) \tag{2.10}$$

for $\alpha \neq 1$, and

$$aX \sim S_\alpha(|a|\, \sigma, sign(a)\beta, a\mu - \frac{2a}{\pi}(log\, |a|)\sigma\beta), \tag{2.11}$$

for $\alpha = 1$.

The above properties are especially useful for modeling the common distribution of various asset returns, under the assumption that the assets are independent and the returns of each obey a stable distribution with one common stability index.

Stable random vectors. A random vector $X = (X_1, ..., X_d)$ is said to be a *stable random vector* in R^d if, for any positive numbers A and B, there is a positive number C and vector $D \in R^d$ such that

$$AX^{(1)} + BX^{(2)} \stackrel{d}{=} CX + D \ , \tag{2.12}$$

where $X^{(1)}$ and $X^{(2)}$ are independent copies of X. The vector is called *strictly stable* if the equation holds with $D = 0$ for any $A > 0$ and $B > 0$.

Characteristic function of an α-stable random vector. [2] Let $X = (X_1, ..., X_d)$ be an α-stable random vector in R^d and let $\Phi_\alpha(\theta) = \Phi_\alpha(\theta_1, ..., \theta_d) = Eexp\{i(\theta, X)\} = Eexp\{i \sum_{k=1}^{d} \theta_k X_k\}$ denote its characteristic function. Then Φ_α is the *joint characteristic function* of the random variables $X_1, ..., X_d$.

The expression of the joint characteristic function of a stable random vector in the following theorem involves an integration over $S_d = \{s : \|s\| = 1\}$, which is the unit sphere in R^d. S_d is a $(d-1)$-dimensional surface.[3]

Theorem 1. Let $0 < \alpha < 2$. Then $X = (X_1, ..., X_d)$ is a stable random vector in R^d if and only if there exists a finite measure Γ on the unit sphere S_d of R^d and a vector μ^0 in R^d such that
(a) if $\alpha \neq 1$ then

$$\Phi_\alpha = exp\left\{-\int_{S_d} |(\theta, s)|^\alpha \left(1 - i\, sign((\theta, s)) \tan\frac{\pi\alpha}{2}\right)\Gamma(ds) + i(\theta, \mu^0)\right\}, \tag{2.13}$$

(b) if $\alpha = 1$ then

$$\Phi_\alpha = exp\left\{-\int_{S_d} |(\theta, s)| \left(1 + i\frac{2}{\pi} sign((\theta, s)) log\, |(\theta, s)|\right)\Gamma(ds) + i(\theta, \mu^0)\right\}. \tag{2.14}$$

The pair (Γ, μ^0) is unique. The vector X in Theorem 1 is said to have a spectral representation (Γ, μ^0). The measure Γ is called the *spectral measure* of the stable random vector X.

Strictly stable random vector. Suppose that $X = (X_1, ..., X_d)$ is an α-stable random vector in R^d with $0 < \alpha \leq 2$. Then X is strictly stable if and only if all its components X_k, $k = 1, ..., d$, are strictly stable random variables.

A necessary and sufficient condition for a strictly stable random vector is that $\mu^0 = 0$, and Γ be a symmetric measure on S_d (i.e. $\Gamma(A) = \Gamma(-A)$), for any Borel set A of S_d.

[2] Samorodnitsky and Taqqu (1994, pp. 65).
[3] S_1 is the two-point set $\{-1, 1\}$ and S_2 is the unit circle.

Theorem 2. [4] X is a *symmetric α-stable vector* in R^d with $0 < \alpha < 2$ if and only if there exists a unique symmetric finite measure Γ on the unit sphere S_d such that

$$Eexp\{i(\theta, X)\} = exp\left\{-\int_{S_d} |(\theta, s)|^\alpha \Gamma(ds)\right\}. \tag{2.15}$$

Γ is the *spectral measure* of the symmetric α-stable random vector.

A symmetric α-stable distribution in R^d is denoted $S\alpha S$. The vector $X = (X_1, ..., X_d)$ is said to be $S\alpha S$ in R^d, and the random variables $X_1, ..., X_d$ are jointly $S\alpha S$.

In order to represent a d-dimensional vector of $S\alpha S$ random variables with common stability index α, the concept of *sub-Gaussian* vectors can be applied. First this property is explained for a single variable: A $S\alpha S$ random variable X can be constructed as $X = A^{1/2}G$, with G as a zero-mean Gaussian random variable, i.e. $G \sim N(0, \sigma^2)$, and with A as an $\alpha/2$-stable random variable totally skewed to the right and independent of G, i.e. $A \sim S_{\alpha/2}\left(\left(cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 1, 0\right)$; A is called $\alpha/2$-stable subordinated.

This result can be extended to the d-dimensional case: Let G be a d-dimensional zero-mean Gaussian random vector, $G = (G_1, ..., G_d)$. Suppose that G is independent of the above-defined $\alpha/2$-stable subordinator A. A d-dimensional $S\alpha S$ vector X is defined by

$$X = (A^{1/2}G_1, ..., A^{1/2}G_d). \tag{2.16}$$

The vector X is called a sub-Gaussian[5] $S\alpha S$ random vector in R^d with underlying Gaussian vector G.

As the covariance for stable distributed random variables with $\alpha < 2$ is always infinite, the concept of sub-Gaussian $S\alpha S$ random vectors can be applied to incorporate the Gaussian dependence structure among stable distributed random variables. Since the Gaussian dependence is easier to calculate, it makes sense to transfer it into the sub-Gaussian case.

One approach toward generating dependent $S\alpha S$ random variables is to use truncated Gaussian covariances from the empirical data. Generating a d-dimensional random vector X is performed by simulating a d-dimensional Gaussian random vector G with correlated elements G_i , $i = 1, ...d$, and an $\alpha/2$-stable random variable A independent of G_i. For details see Rachev, Schwartz and Khindanova (2001).

---

[4] Samorodnitsky and Taqqu (1994, pp. 73).
[5] Samorodnitsky and Taqqu (1994, pp. 77).

2.3 Dependence Among Stable Random Elements

The covariation. The covariance function can only be applied for measuring the dependence among Gaussian random elements ($\alpha = 2$). The *covariation* replaces the covariance for random elements with $1 < \alpha < 2$. Before defining the covariation for jointly $S\alpha S$ random variables, the so-called *signed power* is introduced. The signed power $a^{<p>}$ is defined as

$$a^{<p>} = |a|^p sign(a) = \begin{cases} a^p & if \quad a \geq 0 \\ -|a|^p & if \quad a < 0. \end{cases} \tag{2.17}$$

Definition of the covariation. [6] Let X_1 and X_2 be jointly $S\alpha S$ with $\alpha > 1$ and let Γ be the spectral measure of the random vector (X_1, X_2). Then, the *covariation* of X_1 on X_2 is the real number

$$[X_1, X_2]_\alpha = \int_{S_2} s_1 s_2^{<\alpha-1>} \Gamma(ds). \tag{2.18}$$

Let (X_1, X_2) be jointly $S\alpha S$, $1 < \alpha \leq 2$, and consider the $S\alpha S$ random variable

$$Y = \theta_1 X_1 + \theta_2 X_2, \tag{2.19}$$

where θ_1 and θ_2 are real numbers. Denoting $\sigma(\theta_1, \theta_2)$ as the scale parameter of the random variable Y, there is another definition of the covariation $[X_1, X_2]_\alpha$ equivalent to the one cited in (2.18):

$$[X_1, X_2]_\alpha = \frac{1}{\alpha} \frac{\partial \sigma^\alpha(\theta_1, \theta_2)}{\partial \theta_1} \bigg|_{\theta_1=0, \theta_2=1}. \tag{2.20}$$

Modeling the dependence in a sub-Gaussian random vector. Let $(G_1, ..., G_n)$ be mean-zero jointly Gaussian random variables with covariance $R_{ij} = EG_iG_j$, $i, j = 1, ..., n$, and let $A \sim S_{\alpha/2}\left(\left(cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 1, 0\right)$ be independent of $(G_1, ..., G_n)$. Then, the sub-Gaussian random vector $X = (X_1, ..., X_n)$, with $X_k = A^{1/2}G_k$, $k = 1, ..., n$, $1 < \alpha < 2$, has the following covariations:

$$[X_i, X_j]_\alpha = 2^{-\alpha/2} R_{ij} R_{jj}^{(\alpha-2)/2}. \tag{2.21}$$

Here: if $R_{ii} = R_{jj}$ then $[X_i, X_j]_\alpha = [X_j, X_i]_\alpha$.

However, the covariation in general is not symmetric in its arguments (in contrast to the Gaussian covariance). Thus, there is often:

[6] Samorodnitsky and Taqqu (1994, pp. 87).

$$[X_i, X_j]_\alpha \neq [X_j, X_i]_\alpha. \tag{2.22}$$

The variation. The variance function only exists for Gaussian random elements. It is replaced by the *variation* for random elements with $1 < \alpha < 2$. The variation of a $S\alpha S$ random variable X is defined as

$$Var(X) = [X, X]_\alpha. \tag{2.23}$$

2.4 Studies Of Stable Value at Risk (VaR) And The Normal Case

As stable distributions provide a better fit for financial return data compared to normal fitting, *stable Value at Risk (VaR)* measures are supposed to outperform normal VaR in terms of accuracy. Khindanova, Rachev and Schwartz (1999) have examined stable and normal VaR for market-return data.

A special interest lies in determining the VaR of such financial instruments subject to credit risk for a given time horizon. VaR is a measure for the riskiness of an asset and determines the economic capital required to hold the asset.[7] VaR models seek to measure the maximum loss of value on a given asset or liability over a given time period at a given confidence level (eg. 95%).

Definition of Value at Risk (VaR). The VaR is defined as a threshold regarding the price change of the instrument over the observed time horizon. The return over the time horizon τ is expected to fall below that threshold with a probability of $(1 - c)$. This says that, with a probability of $(1 - c)$, the returns are expected to be less than $-VaR_c$.[8] The VaR is expressed as

$$P[\Delta p(\tau) \leq -VaR_c] = 1 - c, \tag{2.24}$$

with

- $\Delta p(\tau)$: Price change over time horizon τ;
- c: Confidence level of VaR, e.g. 95%;
- The probability that losses will exceed VaR_c is $(1 - c)$.

[7] Saunders (1999, p. 38).
[8] The VaR is defined as a positive number.

Khindanova, Rachev and Schwartz (1999) have performed several empirical tests to measure the differences between empirical and modeled VaR. The VaR as a measure of risk is preferred by financial institutions and regulators, in particular. The quality of modeled VaR is determined by comparing the empirical VaR with the modeled VaR. The fewer the number of cases the empirical VaR exceeds the modeled VaR, the more conservative the VaR model is considered to be.

The results can be summarized as follows:

- Stable modeling produces conservative and more accurate estimates for the VaR at the confidence level 99%, i.e. $c = 0.99$.
- Whereas, stable modeling slightly underestimates the 95% VaR ($c = 0.95$), normal modeling gives more accurate estimates for the 95% level.
- For the 99% VaR, normal modeling leads to overly optimistic forecasts.

The above results were derived from time series of several different market indices.[9] The dominance of stable VaR modeling was also demonstrated for credit return series.[10]

Before introducing a modified one-factor credit model in chapter 3 to describe the returns of credit instruments, a brief summary of known concepts for measuring the performance of financial assets is given. The focal point is specifically the stable case when the limiting distributions of the model's variables and innovations are assumed to follow a stable law.

2.5 Common Performance Measures Under Gaussian And Stable Assumption

In this section a short review of performance measures is presented. First, common performance measures are discussed under the Gaussian assumption for the asset returns. Finally, the section explains how their affiliated models are formulated under the stable assumption.

The performance measures discussed below are commonly used in financial applications. The objective in modeling the returns of a risky financial asset is to find a relationship between risk and return. The risk of an asset can be divided into systematic risk and unsystematic risk. Systematic risk cannot be diversified away and can be linked to external factors. However, the diversifiable risk helps investors to reduce their exposure at no expense. For example, holding a portfolio

[9] Khindanova, Rachev and Schwartz (1999).

[10] Rachev, Schwartz and Khindanova (2001).

consisting of several uncorrelated assets with equal riskiness and equal expected returns allows the investor to reduce risk compared to investing solely in one of these assets. In the following, the Capital Asset Pricing Model (CAPM) is reviewed under both the Gaussian and the stable assumption. The Arbitrage Pricing Theory (APT) is discussed under similar considerations. The Jensen Measure is given as an example for a CAPM-based measure. Regression estimators with multiple risk factors are briefly introduced as well.

2.5.1 The Capital Asset Pricing Model (CAPM)

The Capital Asset Pricing Model (CAPM) assumes a linear relationship between risk and return. Within this framework, there are two types of risk, diversifiable and non-diversifiable risk. The total risk of an asset is measured by the variance of its returns. The systematic (non-diversifiable) risk of a portfolio i describes how sensitive it reacts to market movements. This is measured by the portfolio's β_{im}.

$$\beta_{im} = \frac{cov(R_i, R_m)}{var(R_m)}, \tag{2.25}$$

where R_i is the return on the portfolio i and R_m is the market return.[11]

The relationship between the portfolio's β and the portfolio's expected return is expressed by the Security Market Line (SML). For the SML, only the non-diversifiable risk is relevant:

$$E(R_i) = r_f + (E(R_m) - r_f)\beta_{im} \ , \tag{2.26}$$

where $E(R_i)$ is the portfolio's expected return, r_f is the average risk-free rate, and $E(R_m)$ is the expected return of the market.

Another alternative to describe the expected returns of the portfolio is to plot them against the standard deviation σ_i of the portfolio's returns. The standard deviation accounts for the portfolio's total risk, i.e. systematic and unsystematic risk. The plot is called the Capital Market Line (CML) and follows

$$E(R_i) = r_f + (E(R_m) - r_f)\sigma_i. \tag{2.27}$$

2.5.2 The Stable CAPM

For the CAPM of Sharpe (1964) and Lintner (1965), it was assumed that the asset returns follow a normal distribution. As the CAPM did not prove to be satisfactory when tested empirically, Fama (1970) was the first to introduce symmetrical $S\alpha S$ to model the returns in the CAPM. This work was completed by Gamrowski and

[11] See, for example, Götzenberger, Rachev and Schwartz (2000).

Rachev (1994). In general, Fama's model of asset returns has a more intuitive risk-return relation:

$$R_i = \rho_i + b_i\delta + \epsilon_i, \tag{2.28}$$

where ρ_i is a constant and δ and ϵ_i are independent $S\alpha S$ random variables. Fama used $S\alpha S$ random variables for both the returns and the error term within his stable version for the CAPM. This was due to the following constraints, which the stable variables have to obey:

1. Only symmetric returns are assumed.
2. For computational reasons, all stable distributed variables must have the same index of stability α.

In Fama's CAPM, the equation for the expected returns has the same form as the Sharpe and Lintner model:

$$E(R_i) = \rho_0 + \beta_{im}(E(R_m) - \rho_0), \tag{2.29}$$

where $E(R_i)$ is the expected return of asset i, ρ_0 is the return of the riskless asset, $E(R_m)$ is the expected return of the market portfolio, and β_{im} describes how sensitive the asset i reacts to changes of the market portfolio.

Equation (2.29) can be rewritten as:

$$E(R_i) = \rho_0 + \frac{1}{\sigma(R_m)} \frac{\partial\sigma(R_m)}{\partial(\lambda_{im})} \left[E(R_m) - \rho_0\right], \tag{2.30}$$

where $R_m = \sum_i \lambda_{im} R_i$ represents the return of the market portfolio with $\sum_i \lambda_{im} = 1$.

Thus, with (2.30) and (2.29), the *beta*-coefficient β_{im} is determined by:

$$\beta_{im} = \frac{1}{\sigma(R_m)} \frac{\partial\sigma(R_m)}{\partial(\lambda_{im})} = \frac{1}{\alpha v_\alpha(R_m)} \frac{\partial v_\alpha(R_m)}{\partial\lambda_{im}} . \tag{2.31}$$

In the stable case, there is $\sigma(R_m) = (v_\alpha(R_m))^{1/\alpha}$ with $v_\alpha(R_m)$ as the variation $v_\alpha(R_m) := [R_m, R_m]_\alpha$. Moreover, $\frac{\partial v_\alpha(R_m)}{\partial\lambda_{im}} = \alpha[R_i, R_m]$ with covariation $[R_i, R_m]_\alpha$. [12]

The *stable beta*, β_{im}, can be expressed as:

$$\beta_{im} = \frac{[R_i, R_m]_\alpha}{v_\alpha(R_m)}. \tag{2.32}$$

[12] See the definitions of variation and covariation in Section 2.3.

Recall that with the Gaussian assumption ($\alpha = 2$), there is $\beta_{im} = \frac{cov(R_i,R_m)}{var(R_m)}$ in (2.25). In the stable non-Gaussian case, the *variation* replaces the variance and the *covariation* refers to the covariance.

Fama's case is restricted to the fact that it assumes independence between the factor δ and the innovation ϵ_i. The same results can also be obtained from Ross's mutual fund separation theory.[13]

2.5.3 The Arbitrage Pricing Theory (APT)

The Arbitrage Pricing Theory was developed by Ross (1976). The underlying principle of his theory is the absence of arbitrage. Ross assumes that the realized (ex-post) return of an asset can be described by the (ex-ante) expected return plus changes that are caused by exposure to a number of risk factors and a stochastic error term.

The individual return of an asset i is modeled by

$$R_i = E_i + \sum_{j=1}^{k} \beta_{ij}\delta_j + \epsilon_i \ , \ i = 1, ..., N, \tag{2.33}$$

where E_i is the (ex-ante) expected return, and β_{ij} is the sensitivity of R_i to movements of factor j. N is the total number of assets. The vector δ represents the risk factors j, $j = 1, ..., k$. The expected return E_i of asset i is modeled as

$$E_i = E(R_i) = \rho_0 + \beta_{i1}\rho_1 + ... + \beta_{ik}\rho_k, \ i = 1, ..., N, \tag{2.34}$$

where ρ_l with $l = 1, ..., k$ is the risk premium for exposure to risk factor l.

The challenge of setting up an APT model and its empirical testing is to identify the risk factors σ. This requires the use of multi-factor analysis procedures. Two theories on the APT have evolved from Ross's initial work. The first notion is the asymptotic APT, which assumes a sequence of economies with a growing number of assets. The second is the so-called equilibrium APT, with restrictions imposed on returns and agents' utility functions.

2.5.4 The Stable APT

In this section, a stable version of the APT is discussed. It is related to the asymptotic APT, where a sequence of economies is considered with the n-th economy having n assets. In the asymptotic APT, the returns are generated by a k-factor model[14]

---

[13] Ross (1978).

[14] See, for example, Rachev and Mittnik (2000, pp. 422) and Huberman (1982).

$$R_i^n = E_i^n + \beta_{i1}^n \delta_i^n + ... + \beta_{ik}^n \delta_k^n + \epsilon_i^n. \tag{2.35}$$

Here δ_l^n, $l = 1, ..., k$, are the k factors in the n-th economy. In the stable case, the factors δ and the disturbances ϵ are $S\alpha S$.

$$R^n = E^n + \beta^n \delta^n + \epsilon^n. \tag{2.36}$$

Furthermore, the vector of expected returns E^n can be expressed as

$$E^n = E_0^n + \sum_{j=1}^{k} \beta_j^n \delta_j^n + c^n \ . \tag{2.37}$$

The vector c^n is orthogonal to β^n, and is also chosen in such a way that it is orthogonal to the unit vector e^n; c^n can be interpreted as the arbitrage portfolio which uses no wealth. In case of arbitrage, there would exist a subsequence $\acute{n}$ and with $\acute{n} \to \infty$ the expected excess returns would increase to infinity: $\lim_{\acute{n}\to\infty} E(r^{\acute{n}} c^{\acute{n}}) = +\infty$, with the variation $v_\alpha(r^{\acute{n}}, c^{\acute{n}})$ vanishing: $v_\alpha(r^{\acute{n}}, c^{\acute{n}}) \to 0$.[15]

In case of no arbitrage for the economy with n assets, $n = 1, 2, ...$, there exists an E_0^n, a sequence γ_j^n, and an A, such that

$$\sum_{i=1}^{n} \left| E_i^n - E_0^n - \sum_{j=1}^{k} (\beta_{ij}^n \gamma_j^n) \right|^\alpha \leq A. \tag{2.38}$$

The relationship (2.38) says that, in large economies, the mean returns are linearly correlated with the economy's risk factors.[16]

In the Gaussian case, the β_{ij}^n are represented by the covariances between the asset returns and the risk factors. In the α-stable case, covariations are applied instead of covariances.

2.5.5 The Jensen Measure

One of the most widely accepted CAPM-based performance measures is the *Jensen Measure*, which describes the excess returns of a portfolio by applying a linear regression over the market excess returns:[17]

$$r_{pt} = \alpha_p + \beta_p r_{mt} + u_{pt}. \tag{2.39}$$

The β_p measures the systematic risk of the portfolio, indicating how sensitive the asset or portfolio reacts to movements of the market. If α_p is significantly

[15] See Götzenberger, Rachev and Schwartz (2000).
[16] Rachev and Mittnik (2000).
[17] Jensen (1968).

greater than zero, the modeled portfolio is expected to outperform the market portfolio. Jensen's Multiperiod Model is based on the SML as benchmark (see equation (2.26)).

The drawbacks of the CAPM were discussed by Roll (1977), who argued that the true market portfolio would not be observable. Therefore, it would be impossible to determine the benchmark. Furthermore, Roll found that securities would plot on the SML if and only if the market portfolio is efficient.

Jensen's α_p does not give information as to whether an asset or portfolio shows superior performance. Moreover, as Roll pointed out, if the market portfolio could be observed, and if it were efficient, the securities would all plot on the security market line (SML). This would mean that superior or inferior performance could not be possible. Significant positive or negative values for α_p could not occur.

In 1978, Roll tested the CAPM model with three different benchmarks for calculating the βs. He received three different rankings regarding the performance of the observed assets.[18]

2.5.6 Regression Estimators

In the following, possible estimators for the determination of regression parameters are discussed - under the assumption that the random variables of the regression follow a stable law with $\alpha < 2$. As an example, the focus is on the stable case of Jensen's Measure (2.39) as risk-return relation:

$$r_{pt} = \alpha_p + \beta_p r_{mt} + u_{pt}, \tag{2.40}$$

where r_{mt} is the single risk factor. Assume the distributions of the risk factor and the innovation to be $S\alpha S$-stable with $\alpha < 2$. The OLSE [19] to determine $\hat{\beta}_p$ is no longer BLUE[20]. OLSE is still unbiased but is no more efficient. However, OLSE can still be used:

$$\hat{\beta}_p = \frac{\sum_t r_{mt} r_{pt}}{\sum_t r_{mt}^2} \ , \hat{\alpha}_p = \bar{r}_p - \bar{r}_m \hat{\beta}_p.$$

For multiple regressions with several risk factors the OLSE can also be applied. The case that all risk factors are independent of each other is very unlikely. Thus, the focus is now on the dependent case. Constructing a sub-Gaussian $S\alpha S$ vector $X \stackrel{d}{=} A^{1/2} G$, the dependence among the components of X can be described as

$$E(X_n | X_1, ..., X_{n-1}) = c_1 X_1 + ... + c_{n-1} X_{n-1}, \tag{2.41}$$

[18] Roll (1978).
[19] OLSE = Ordinary Least Squares Estimator.
[20] BLUE = Best Linear Unbiased Estimator.

where

$$c_i = [X_n, X_i]_\alpha / v_\alpha(X_i) = cov(G_n, G_i)/var(G_i), \tag{2.42}$$

and $G = (G_1, ..., G_n)$ is the underlying mean-zero Gaussian vector of X. Therefore, it is obvious that the OLS estimates can be applied.

Jensen's measure containing several $(k \geq 2)$ risk factors is described as

$$r_{pt} = \alpha_p + \beta_1 r_{f1,t} + ... + \beta_k r_{fk,t} + u_{pt}. \tag{2.43}$$

With vector notation, this is written in the form

$$r_{pt} = r_f \beta + u_{pt}, \tag{2.44}$$

where $\beta^T = (\alpha_p, \beta_1, ..., \beta_k)$ and $r_f = (1, r_{f1}, r_{f2}, ..., r_{fk})$ is a $t \times (k+1)$-matrix.

The OLSE $\hat{\beta}$ is obtained by

$$\hat{\beta} = (r_f^T, r_f)^{-1} r_f^T r_p. \tag{2.45}$$

In this chapter, the stable distribution and its properties have been introduced to the reader. The modeling of dependent asset returns is performed via stable random vectors. As the covariation is highly complex to be used, the dependence between stable random vectors should be easier modeled by a so-called sub-Gaussian vector. Variation and VaR are measures that indicate the riskiness of an asset. Financial returns exhibit peakedness and heavy-tailedness, which makes stable VaR superior to its Gaussian counterpart.

The final section has presented an overview on commonly known performance measures and their definition under the stable assumption.

The properties of stable distributed random variables (e.g. sum, multiplication) and vectors help to model the behavior of a portfolio consisting of various assets whose returns follow a stable law.

Chapter 3

Stable Modeling In Credit Risk

This chapter starts with a brief review of recent advances in stable modeling of credit risk discussing the approach taken by Rachev, Schwartz and Khindanova (2001).

Then, a new approach based on a modification of their model is presented. The modification is done by a change in the definition of bond returns. This approach can be implemented more easily, also because its data requirements are less problematic. The modified model is built both for the Gaussian and for the stable assumption. Furthermore, the case when the returns of different credit instruments are assumed to be independent and the case when the returns are assumed to be dependent are both covered.

In order to illustrate the effects of the different assumptions (stable vs. Gaussian, dependent vs. independent), an empirical example is set out. A portfolio consisting of two corporate bonds is chosen and its daily VaR is calculated for each combination of the assumptions.

3.1 Recent Advances

Academics and practitioners[1] have examined the application of stable distributions for modeling asset returns. As it is well documented in the literature on empirical finance[2], changes in value of a financial asset are heavy-tailed and peaked, whereas the mass of the commonly used normal distribution is located around its center. For this reason, the normal assumption fails to model crashes and strong upturns in financial markets.

Recent research has also examined daily returns of assets subject to credit risk.

[1]See the work of Mandelbrot (1963), Fama (1965a), and Fama (1965b).

[2]See, for example, Rachev and Mittnik (2000).

Studies[3] found that credit returns are also peaked and heavy-tailed. Moreover, they turned out to be skewed.

Rachev, Schwartz and Khindanova (2001) suggested the application of stable distributions for credit instruments to meet those properties. As explained above, for stable distributions, the peakedness and the heavy tails are determined by the stability index α, whereas the parameter β is responsible for skewness or asymmetry.

3.2 A One-Factor Model For Stable Credit Returns

This section reviews the credit model derived by Rachev, Schwartz and Khindanova (2001).

In their model for credit returns, Rachev, Schwartz, and Khindanova assumed a linear relationship between the returns of a risky credit instrument and the returns of a comparable risk-free credit instrument.

For such a credit instrument i, the returns are described by

$$R_i = a_i + b_i Y_i + U_i, \tag{3.1}$$

where

- R_i are the log returns of an asset i that is subject to credit risk.
- Y_i are the log returns of a risk-free asset.
- U_i is the disturbance. It represents the spread or the premium for the credit risk.
- a_i and b_i are constants which are obtained by ordinary least squares (OLS) estimation.

In this linear model, the returns of both the risky (R_i) and the risk-free (Y_i) credit instrument are assumed to follow a strictly stable law. Moreover, the disturbance term (U_i) is a strictly stable random variable:

- $U_i \sim S_\alpha(\sigma_\alpha, \beta_\alpha, \mu_\alpha)$, $1 < \alpha < 2$;
- $Y_i \sim S_\gamma(\sigma_\gamma, \beta_\gamma, \mu_\gamma)$, $1 < \gamma < 2$.

[3] Federal-Reserve-System-Task-Force (1998), Basle-Committee (1999).

For credit instruments, the log return $R_{i,t}$ at time t is defined as

$$R_{i,t} = log\left(\frac{P_{i,t,T}}{P_{i,t-1,T-1}}\right), \quad (3.2)$$

where $P_{i,t,T}$ is the price of an instrument i subject to credit risk with maturity date T evaluated at time t. The log returns of the riskless asset $Y_{i,t}$ are determined by

$$Y_{i,t} = log\left(\frac{B_{i,t,T}}{B_{i,t-1,T-1}}\right), \quad (3.3)$$

with $B_{i,t,T}$ as the price of the risk-free asset with maturity date T evaluated at time t. This means that all prices used for the calculation of the returns are determined on the basis of constant *time to maturity.* Therefore, the time series of log returns (both $Y_{i,t}$ and $R_{i,t}$) is calculated such that the time to maturity is the same for all t.

It must be noted that $Y_{i,t}$ and $R_{i,t}$ are not directly observable for individual bonds whose market price movements are recorded on a daily basis. The prices $B_{i,t,T}, B_{i,t-1,T-1}, B_{i,t-2,T-2}, ...$ are calculated from the yield curve of riskless treasury bonds and $P_{i,t,T}, P_{i,t-1,T-1}, P_{i,t-2,T-2}, ...$ are derived from a yield curve generated from risky bonds representing a similar level of credit risk (e.g. having equal credit ratings). Such an approach enables the risk manager to deal with constant time to maturity. This is crucial, because for the prices of individual bonds, time to maturity decreases with increasing time t.

The effect of changing time to maturity on credit returns can be demonstrated by a small example with two riskless zero bonds: one with a time to maturity of one year, the other with a time to maturity of two years. Furthermore, the term structure is assumed to be flat, and therefore, both securities have equal yields. If the yield of both increases by the same percentage, then the price of the two-year bond reacts more sensitively compared to the one-year bond.

However, the approach of modeling the returns as set out in (3.2) and (3.3) is very difficult to implement in practice. Historical data of daily yield curves is available for treasury bonds (for the $B_{i,t,T}, B_{i,t-1,T-1}, B_{i,t-2,T-2}, ...$), but it is practically impossible to observe a time series of prices $P_{t,T}, P_{t-1,T-1}, P_{t-2,T-2}, ...$ for an individual bond. Therefore, a number of different credit risk categories has to be defined first and individual corporate bonds have to be assigned to them.[4] The prices of numerous bonds assigned to the same risk category are then taken to generate the corresponding yield curve.[5]

[4]For example, the rating grades assigned by Standard & Poor's or Moody's may be employed to define the risk categories.

[5]For example, see McCulloch (1971) and McCulloch (1975).

In order to avoid such difficulties, a more practical way to define the credit returns has to be found. Obviously, a risk manager would prefer to deal with the observed real prices of a bond to fit a model, rather than deriving prices from yield curves that first have to be generated. Moreover, each yield curve only represents an average credit risk level. The approach proposed in the following section allows to determine the individual credit risk of the bond analyzed.

3.3 A New Approach For The Returns

Having historical daily yield curve data of treasury bonds available, this allows to construct historical daily prices for any treasury bond with given coupon, coupon dates, and maturity. Thus, a corresponding riskless[6] treasury bond i with identical specifications can be generated for each risky corporate bond i. In this case, the return $R_{i,t}$ of a risky corporate bond is defined as its actual (observable) daily price movement:

$$R_{i,t} = log\left(\frac{P_{i,t,T}}{P_{i,t-1,T}}\right). \tag{3.4}$$

Here, time to maturity is no longer kept fixed. The return $R_{i,t}$ is that of an individual bond i with fixed maturity date T. The riskless returns $Y_{i,t}$ are defined analogously:

$$Y_{i,t} = log\left(\frac{B_{i,t,T}}{B_{i,t-1,T}}\right). \tag{3.5}$$

This riskless bond i has the same specifications (maturity, coupon, coupon dates), as the risky bond i.

With this new approach, the original linear risk-return relation $R_i = a_i + b_i Y_i + U_i$ remains, but its components R_i, Y_i, and U_i now have a different meaning. R_i and Y_i are individual bond returns, and the disturbance U_i incorporates both credit spread and the risk of time to maturity.

For the empirical examinations in this chapter, the model with the returns defined in (3.4) and (3.5) was utilized. In the following, a brief summary of the advantages and disadvantages of both approaches (definitions (3.2) and (3.3) vs. definitions (3.4) and (3.5) is presented:

- The model whose returns are defined by the equations (3.2) and (3.3) abandons the problem to deal with changing time to maturity. This is its key advantage.

[6] "Riskless" in this context refers to "free of credit risk".

- The drawback of such an approach is that yield curves have to be modeled first for a number of different risk levels (e.g. corporate credit ratings) and for the risk-free (treasury) bonds.

- After fitting the parameters a and b of equation (3.1), future scenarios are simulated for each yield curve. Such a framework enables simulation of future daily returns for each time to maturity. Finally, the simulated yield curves can be employed to derive the simulated future returns of individual corporate bonds. This procedure makes the approach of (3.2) and (3.3) very complex.

- With the model defined by the returns in (3.4) and (3.5) there is no need to construct yield curves for a number of different credit risk levels (of risky corporate bonds), nor is it necessary to simulate future representations of such yield curves by applying a term structure model. In fact, this model allows future returns of individual bonds to be simulated directly from their fitted distributions of Y_i and U_i.

- It was expected that for real bonds with decreasing time to maturity, there would be an impact on the return distribution when time is moving towards maturity. However, performing empirical testing with numerous sample bonds, a decreasing time to maturity does not seem to have a noticeable effect on the distribution of the credit returns defined by (3.4) and (3.5). The test compared the distribution of the returns for different intervals of the available time series. A systematic difference in the return distribution between intervals lying more distant to maturity point and intervals closer to maturity point could not be observed empirically.

This is why the approach based on the definitions in (3.4) and (3.5) is selected.

Thus, the first key advantage of the chosen approach is the ability to work with the actual historical price data and spread information of the individual bonds, instead of generating yield curves, each for a certain risk grade. Such yield curves only represent the average of a risk grade. Studies (e.g. Beck, 2001) found that, in some cases, a higher rated bond may even have a larger credit spread than bonds with a lower rating grade. This is due to the fact that the range of credit spreads within a given rating grade may be relatively wide and that the spread ranges of neighboring grades usually overlap. This effect has also been illustrated by (Kealhofer, Kwok and Weng, 1998). [7] In the past other researchers, e.g. such

[7] A reason for this effect could be that the market values the creditworthiness of an issuer differently than the rating agencies do. Sometimes the market anticipates a change in the credit quality of an issuer before the rating agencies react.

as (Katz, 1974) or (Fisher, 1959), have analyzed risk premiums of bonds and their relation to rating grades or rating grade changes.

The construction of a yield curve for a given credit rating usually requires data from a large number of bonds with various issuers. The yield curve of a single issuer is calculable even only for large corporations with a large quantity of issued bonds.

The second key advantage of the approach selected is that it can easily be implemented into practice:

- The model does not require the construction of yield curves for a number of risk levels (of risky bonds), nor does it necessitate the simulation of future representations of the yield curves by applying a complex term structure model.
- The model enables direct simulation of the future returns of individual corporate bonds by generating representations of Y_i and U_i according to their fitted distributions.

However, it has to be noted that the fitted distribution of U_i not only accounts for the credit risk but also for liquidity risk. Liquidity will not be further considered for the credit model. Although, the following section provides a brief excursus on this topic.

3.4 Excursus: Liquidity Risk

Another source of risk that influences the movement of bond prices is the liquidity present in the market. This section is an excursus and briefly addresses the issue of liquidity even though the credit model does not allow for it.

Liquidity is defined as the speed and ease at which one can trade. A market is liquid if one can trade a large quantity shortly after the desire to trade at a price near the prices of the trades prior and after the desired trade.[8]

The issue of liquidity should just be touched on here, it will not be treated extensively.

As there are no perfectly liquid bond markets, in the model (3.1), the disturbance term U_i also accounts for liquidity risk. Periods of serious illiquidity are often visible and the credit spread of bonds is influenced by the changing liquidity of the market.[9]

When a seller is not able to find a buyer for an asset at a fair price due to lack of liquidity, this imposes a negative component on the daily price change.

[8] Huberman and Halka (1999).

[9] Chordia, Roll and Subrahmanyan (2000a).

Considering the given credit model, a way is needed to separate the disturbance term U_i into a component driven by credit risk and into another component that accounts for liquidity risk.

Although liquidity varies from security to security, it makes sense to search for a common measure for liquidity which will describe the liquidity of the whole bond market. As liquidity is not observable on its own, a proxy has to be applied to describe it. Huberman and Halka (1999) tested four proxies of liquidity: spread, spread/price ratio, quantity depth, and dollar depth. When examining the average time series of liquidity proxies for two mutually exclusive subsets of stocks, Huberman and Halka found that the innovations of the proxies' time series are positively correlated. This indicates a common liquidity shock. Therefore, it is reasonable to assume a common liquidity component for the market. Chordia, Roll and Subrahmanyan (2000b) estimate a market model that regresses daily percental changes of liquidity variables for individual stocks on the market averages of the same variables. Their resulting betas were positive for 85% of all stocks in their sample. And 42% of the sample showed positive betas that were significant at the 95% level.

I might be useful to define a common liquidity proxy for the whole bond market or a subset of the market and then to isolate the liquidity component from the U in model (3.1). However, as mentioned, this issue shall not be pursued here.

In order to determine the fraction of daily returns that is caused by changes in liquidity, it would be suitable to have a credit instrument whose credit quality remains nearly constant over time. Selecting an index over a number of corporate bonds with a given credit rating would be one alternative. For example, given an index of BBB-rated corporate bonds, its price changes are, aside from changes of the riskless interest rate, largely due to changes of liquidity within this particular bond market.[10]

In less liquid bond markets (e.g. with small issues, small issuers), strong mismatches of supply and demand can occur, causing effects on the credit spread, but the company's actual credit risk remains unaffected by this. Chordia, Roll and Subrahmanyan (2000a) conducted a market study in order to identify the drivers of liquidity and trading activity. They found interesting regularities which determined liquidity and trading activity. Liquidity was represented by quoted and effective spread as well as market depth in this study. Trading activity was represented by measures such as trading volume and the number of daily transactions. Over the observed time period (1988 - 1998) trading activity showed greater variances (average absolute change between 10% and 14%), compared to liquidity

[10] Of course, there may be influences from changes in overall credit quality among the BBB-rated bonds. For example, during an economic recession the default probability of all BBB-rated bonds can on average increase. However, the prevalent impact on the spread of the BBB-index is supposed to be caused by changes in liquidity.

(average absolute change of about 2%).

According to Demsetz (1968), who assumed that liquidity depends on dealers' financing costs and inventory risk, the initially chosen explanatory variables for liquidity in a study performed by Chordia, Roll and Subrahmanyan (2000a) were short-/long-term interest rates, default spreads (generated from market volatility), and contemporaneous moves. Moreover, market-wide changes in liquidity were likely to occur immediately before official announcement events, e.g. key figures on the state of the economy. Furthermore, Chordia et al. assumed that liquidity varies during the week and around holidays due to variations in trading cost.

For most series, they found an AR process of 5th order which indicates a weekly cycling. Equity market performance is also seen as an important factor, as recent market movements will change investors future expectations and trigger restructuring of portfolios.

Summing up their results, Chordia et al. concluded that trading activity responds to short-term interest rates, term spread (the spread between a one-year and ten-year treasury bill), equity market returns, and recent market volatility.

The liquidity (measured by the spreads), responds to equity, recent market activity, and recent market volatility.

It would take further examinations to determine how these results could be transferred to the previously introduced model for credit returns.

3.5 Credit Risk Evaluation For Single Assets

In order to obtain the VaR for a bond i over a time horizon of one period, the following steps are performed:

- A corresponding risk-free treasury bond with equal maturity, coupon, and coupon dates is created.
- The estimates for a_i and b_i are calculated with OLSE.

As set out in Rachev, Schwartz and Khindanova (2001), the estimates are given by

$$\hat{a_i} = \frac{\sum_{t=1}^{T} Y_{it}^2 \sum_{t=1}^{T} R_{it} - \sum_{t=1}^{T} Y_{it} \sum_{t=1}^{T} R_{it} Y_{it}}{T \sum_{t=1}^{T} Y_{it}^2 - (\sum_{t=1}^{T} Y_{it})^2} \tag{3.6}$$

$$\hat{b_i} = \frac{T \sum_{t=1}^{T} R_{it} Y_{it} - \sum_{t=1}^{T} Y_{it} \sum_{t=1}^{T} R_{it}}{T \sum_{t=1}^{T} Y_{it}^2 - (\sum_{t=1}^{T} Y_{it})^2} \tag{3.7}$$

where $i = 1, ..., N$; $t = 1, ..., T$.

With the estimates $\hat{a}_i$ and $\hat{b}_i$, the residuals $\hat{U}_i$ are obtained.

$$\hat{U}_i = R_i - \hat{a}_i - \hat{b}_i Y_i. \tag{3.8}$$

- Finally, a stable fit of $\hat{U}_i$ and Y_i is performed.
- In order to calculate the VaR of asset i for one period, representations of $R_i = a_i + b_i Y_i + U_i$

are simulated. In this case, a sample size of 1000 simulations is chosen.

3.6 A Portfolio Model With Independent Credit Returns

Assume there are n different credit instruments i (bonds) in a portfolio and let v_i be the weight of security i within the portfolio.[11] The return of the portfolio is given by

$$R_p = \sum_{i=1}^{n} v_i R_i \ , \tag{3.9}$$

with

$$R_p = \sum_{i=1}^{n} v_i \left(a_i + b_i Y_i + U_i\right) = \sum_{i=1}^{n} v_i a_i + \sum_{i=1}^{n} v_i b_i Y_i + \sum_{i=1}^{n} v_i U_i, \tag{3.10}$$

and

$$\sum_{i=1}^{n} v_i = 1. \tag{3.11}$$

R_p can be expressed by

$$R_p = \sum_{i=1}^{n} v_i a_i + Y_p + U_p, \tag{3.12}$$

with Y_P and U_p given by

$$Y_p = \sum_{i=1}^{n} v_i b_i Y_i \tag{3.13}$$

[11] v_i may also be negative if short-selling is permitted.

and

$$U_p = \sum_{i=1}^{n} v_i U_i. \tag{3.14}$$

The constant a_p of the total portfolio is

$$a_p = \sum_{i=1}^{n} v_i a_i. \tag{3.15}$$

Assuming the R_i are driven by independent α-stable distributions, this also means that both the U_i and the Y_i, $i = 1...n$, are independent of each other. Assume further that both the U_i and the Y_i, $i = 1...n$, are characterized by a common index of stability (α for the U_i, γ for the Y_i). A common stability index allows an easy analytical solution for the parameters of the distributions for U_p and Y_p. For the properties of stable distributions, the reader is referred to section 2.2 or Samorodnitsky and Taqqu (1994, chapt. 1).

The common index of stability is calculated as an average from the stability indices of the distributions of the individual U_i and Y_i, weighted according to formula (3.9):

$$\alpha = \frac{\sum_{i=1}^{n} |v_i| \alpha_i}{\sum_{i=1}^{n} |v_i|} \tag{3.16}$$

and

$$\gamma = \frac{\sum_{i=1}^{n} |v_i| \gamma_i}{\sum_{i=1}^{n} |v_i|}. \tag{3.17}$$

With the common stability index, the parameters β, σ, μ first have to be reestimated for the individual U_i and Y_i.

The assumption of independent returns provides an analytical solution for the portfolio's U_p and Y_p.

The parameters of U_p and Y_p are then determined by the following expressions:

$$\sigma_{U_p} = \left[\sum_{i=1}^{n} (|v_i| \, \sigma_{U_i})^{\alpha} \right]^{1/\alpha} , \tag{3.18}$$

$$\beta_{U_p} = \frac{\sum_{i=1}^{n} [sign(v_i) \beta_{U_i} (|v_i| \, \sigma_{U_i})^{\alpha}]}{\sum_{i=1}^{n} (|v_i| \, \sigma_{U_i})^{\alpha}} , \tag{3.19}$$

$$\mu_{U_p} = \sum_{i=1}^{n} v_i \mu_{U_i} , \tag{3.20}$$

$$\sigma_{Y_p} = \left[\sum_{i=1}^{n} (\left| v_i \hat{b}_i \right| \sigma_{Y_i})^{\gamma} \right]^{1/\gamma} , \tag{3.21}$$

$$\beta_{Y_p} = \frac{\sum_{i=1}^{n}[sign(v_i\hat{b}_i)\beta_{Y_i}(\left|v_i\hat{b}_i\right|\sigma_{Y_i})^\gamma]}{\sum_{i=1}^{n}(\left|v_i\hat{b}_i\right|\sigma_{Y_i})^\gamma} , \tag{3.22}$$

$$\mu_{Y_p} = \sum_{i=1}^{n} v_i\mu_{Y_i} . \tag{3.23}$$

The portfolio's returns R_p are given by (3.12).

3.7 A Stable Portfolio Model With Dependent Credit Returns

This section introduces a solution for modeling the dependence between credit returns and integrating the skewness-property of their distributions.

Each variable U_i and Y_i is split into a dependent, symmetric and into an independent, skewed component. Both components are independent of each other:

$$U_i = U_i^{(1)} + U_i^{(2)} , \tag{3.24}$$

$$Y_i = Y_i^{(1)} + Y_i^{(2)} . \tag{3.25}$$

The example of U_i demonstrates the derivation of the parameters for the two independent components. Both components are defined as having identical stability indices:

$$U_i^{(1)} \sim S_\alpha(\sigma_1, \beta_1, 0) , \tag{3.26}$$

$$U_i^{(2)} \sim S_\alpha(\sigma_2, \beta_2, 0) . \tag{3.27}$$

Because of the independence of $U_i^{(1)}$ and $U_i^{(2)}$, the parameters' values of U_i are calculated as follows:

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \tag{3.28}$$

$$\beta = \frac{\beta_1\sigma_1^\alpha + \beta_2\sigma_2^\alpha}{(\sigma_1^\alpha + \sigma_2^\alpha)}. \tag{3.29}$$

$U_i^{(1)}$ is symmetric, therefore $\beta_1 = 0$. Equal values for the scale parameters σ_1 and σ_2 are set: $\sigma_1 = \sigma_2 = \sigma^*$.

Thus, the parameters of U_i are:

$$\sigma = 2^{1/\alpha}\sigma^*, \tag{3.30}$$

$$\beta = \frac{1}{2}\beta_2. \tag{3.31}$$

Summing up the results for the parameters: $\sigma_1 = \sigma_2 = \sigma^* = 2^{-1/\alpha}\sigma$, $\beta_2 = 2\beta$ (β_2 is for the skewed component $U_i^{(2)}$), and $\beta_1 = 0$ (β_1 is for the symmetrical component $U_i^{(1)}$),

$$U_i^{(1)} \sim S_\alpha(2^{-1/\alpha}\sigma, 0, 0), \tag{3.32}$$

$$U_i^{(2)} \sim S_\alpha(2^{-1/\alpha}\sigma, 2\beta, 0). \tag{3.33}$$

Analogously, Y_i is split into $Y_i^{(1)} + Y_i^{(2)}$, and their parameters are obtained the same way.

The return of the credit instrument i is then given by

$$R_{i,t} = a + b(Y_{i,t}^{(1)} + Y_{i,t}^{(2)}) + (U_{i,t}^{(1)} + U_{i,t}^{(2)}). \tag{3.34}$$

The symmetric components $Y_{i,t}^{(1)}$ and $U_{i,t}^{(1)}$ are used to incorporate the dependence among the n assets. The dependence structure of the $S\alpha S$[12] vectors $(U_1^{(1)}, U_2^{(1)}, ..., U_n^{(1)})$ and $(Y_1^{(1)}, Y_2^{(1)}, ..., Y_n^{(1)})$ is modeled by representation as sub-Gaussian vectors. Thus, $(U_1^{(1)}, U_2^{(1)}, ..., U_n^{(1)})$ is represented as

$$(U_1^{(1)}, U_2^{(1)}, ..., U_n^{(1)}) \sim (A^{1/2}G_1, A^{1/2}G_2, ..., A^{1/2}G_n), \tag{3.35}$$

where A is a totally skewed $\alpha/2$-stable random variable with $A \sim S_{\alpha/2}((cos\frac{\Pi\alpha}{4})^{2/\alpha}, 1, 0)$ and $G = (G_1, G_2, ..., G_n)$ is an n-dimensional Gaussian zero mean random vector. Let $R_{ij} = EG_iG_j$, $i, j = 1...n$, be the covariances within the vector $G = (G_1, G_2, ..., G_n)$. Then $(U_1^{(1)}, U_2^{(1)}, ..., U_n^{(1)})$ is generated by simulating a representation of the Gaussian vector G with correlated elements $G_1, G_2, ..., G_n$ and an independent representation of the $\alpha/2$-stable random variable A.[13]

The generation of vector $(Y_1^{(1)}, Y_2^{(1)}, ..., Y_n^{(1)})$ is performed analogously.

The results presented in this chapter so far can be summed up as follows:

- A modification of the model by Rachev, Khindanova, and Schwartz has been introduced. Changing the definition of bond returns makes the model more practical and easier to implement.
- The dependence is modeled via a sub-Gaussian vector which is a practical way to incorporate stochastic dependence measured by Gaussian correlations into the stable case.

[12] A $S\alpha S$ vector is a symmetrically stable random vector.

[13] There are various ways to model the dependence. For example, see Rachev, Schwartz and Khindanova (2001).

In the following, the various model assumptions (stable vs. Gaussian, dependent vs. independent) have to be evaluated empirically. This is presented in the final section of this chapter.

3.8 Comparison Of Empirical Results

3.8.1 The Observed Portfolio Data

For the sample portfolio, two corporate bonds have been selected from the US corporate bond market. Both bonds pay coupons twice a year. Historical prices were obtained from Bloomberg[14] for the past four years (March 14 1996 - March 13 2000). According to their credit ratings, the bonds exhibit considerable credit risk. For the given portfolio, there is one unit of each security. Both have a nominal value of US $ 100 (see table 3.1).

	Corporation	Coupon	Rating (S&P / Moodys)	Maturity
1.	Pennzoil (Bond 1)	10.25	BBB+ / Baa2	11/05
2.	United Airlines (Bond 2)	9.0	BB+ / Baa2	12/03

Table 3.1: Bonds selected for sample portfolio.

3.8.2 Generating Comparable Risk-Free Bonds From The Yield Curve

First, the daily returns are calculated from the historical market prices for the period March 14 1996 - March 13 2000. Then, for each bond, a corresponding riskless bond is generated in order to derive the values of the Y_i for the same time period. The corresponding riskless bond has the same specifications (maturity, coupon, coupon date) as the risky corporate bond. The history of daily prices of these artificial treasury bonds is calculated from the daily treasury yield curves. The treasury-yield curve for each day is approximated by prices of 9 risk-free zero bonds with maturities: 0.25, 0.5, 1, 2, 3, 4, 5, 7, 10 years. These 9 points are interpolated by a natural cubic spline algorithm.[15]

With the daily yield curves obtained, historical prices for the artificial treasury bonds are generated. Their daily returns are calculated according to (3.5).

Next, the linear regressions to estimate the parameters a and b of the equations $R_i = a_i + b_i Y_i + U_i$ are performed. $\hat{a}_i$ and $\hat{b}_i$ are OLS estimates (see equations (3.6) and (3.7) and table 3.2).

[14] Bloomberg Information System, Corporate Bonds Section.
[15] Burden and Faires (1997).

Corporation	$\hat{a}_i$	$\hat{b}_i$
Pennzoil (Bond 1)	0.0000	0.9262
United Airlines (Bond 2)	0.0000	0.9878

Table 3.2: Estimates for a and b.

With the values $\hat{b}$ and $\hat{a}$, the estimates for the disturbances U_i can be calculated: $\hat{U}_i = R_i - \hat{a} - \hat{b}Y_i$. Now, both stable and normal fit are applied to the empirical distributions of R_i, Y_i, and $\hat{U}_i$.

During the available sample period from 1996 to 2000, time to maturity for the observed bonds drops by 41% and 52%. The question arises whether this has a systematic effect on the estimated parameters of the distributions of the Y_i over time. However, in the given case, empirical analysis found no evidence to support this.

3.8.3 Fitting The Empirical Time Series For R_i, Y_i, and $\hat{U}_i$

For the stable fit, maximum likelihood estimation (MLE) is applied applied to obtain the four parameters of the distribution. The stable densities were approximated via Fast Fourier Transformation.[16] The procedure was implemented with Matlab 5.3.

The parameters of the stable and Gaussian distributions fitted for the R_i, Y_i, and U_i are shown in tables 3.3, 3.4, and 3.5:[17]

	stable				normal	
Corporation	alpha	beta	sigma	mu	mean	sd
Pennzoil (Bond 1)	1.5451	-0.0690	0.0019	0.0000	-0.0001	0.0041
United Airlines (Bond 2)	1.5199	-0.0744	0.00164	0.0000	-0.0001	0.0035

Table 3.3: Parameters for R fitted with stable and normal distribution.

3.8.4 VaR-Results For The Independence Assumption

The assumption of independence between the bonds in the portfolio leads to the application of the equations in section 3.6. The stable fit for both the Y_i and the U_i is based on a common stability index: $\alpha = 1.10$ is selected for the U_i and $\gamma = 1.32$

[16] For example, see Rachev and Mittnik (2000, pp. 120).

[17] "sd" denotes the standard deviation.

	stable				normal	
Corporation	alpha	beta	sigma	mu	mean	sd
Pennzoil (Bond 1)	1.3639	-0.0297	0.0012	0.0000	-0.0001	0.0027
United Airlines (Bond 2)	1.2811	0.0062	0.0009	0.0000	-0.0001	0.0022

Table 3.4: Parameters for Y fitted with stable and normal distribution.

	stable				normal	
Corporation	alpha	beta	sigma	mu	mean	sd
Pennzoil (Bond 1)	1.0348	-0.0247	0.0006	0.0000	0.0001	0.0027
United Airlines (Bond 2)	1.1663	0.0117	0.0008	0.0000	0.0000	0.0032

Table 3.5: Parameters for the disturbance U fitted with stable and normal distribution.

for the Y_i. Re-estimating the parameters by a stable fit applying common stability indices, one yields the results presented in tables 3.6 and 3.7.

	stable			
Corporation	alpha	beta	sigma	mu
Pennzoil (Bond 1)	1.1000	-0.0047	0.0007	0.0000
United Airlines (Bond 2)	1.1000	0.0828	0.0011	0.0000

Table 3.6: Parameters for the disturbance U fitted with stable and normal distribution assuming $\alpha = 1.10$.

	stable			
Corporation	alpha	beta	sigma	mu
Pennzoil (Bond 1)	1.3200	0.0089	0.0013	0.0001
United Airlines (Bond 2)	1.3200	-0.0430	0.0010	0.0000

Table 3.7: Parameters for Y fitted with stable and normal distribution assuming $\gamma = 1.32$.

The parameters of the portfolio's U_p and Y_p are given by

$$U_p = v_1 U_1 + v_2 U_2 \quad \text{and} \quad Y_p = v_1 \hat{b}_1 Y_1 + v_2 \hat{b}_1 Y_2, \tag{3.36}$$

are determined by the relationships presented in section 3.6. With $v_1 = v_2 = 0.5$, U_p and Y_p are

$$U_p = 0.5U_1 + 0.5U_2 \tag{3.37}$$
and
$$Y_p = 0.5\hat{b}_1 Y_1 + 0.5\hat{b}_1 Y_2.$$

The results for the parameters of U_p and Y_p are printed in table 3.8. Their calculation is performed according to the equations (3.18) - (3.23).

	stable			
	alpha	beta	sigma	mu
Y_p	1.3200	-0.0137	0.0009	0.0000
U_p	1.1000	0.0497	0.0008	0.0000

Table 3.8: Stable parameters for portfolio U_p and Y_p.

The resulting equation describes the portfolio's returns:

$$R_p = 0.5(\hat{a}_1 + \hat{a}_2) + Y_p + U_p = 0.0000 + Y_p + U_p\ . \tag{3.38}$$

Based on this, 1000 daily returns R_p are simulated by generating Y_p and U_p. This yields the daily VaR of the portfolio. For the stable model with independence assumption, a one-day VaR of 0.67% is obtained at the 95% level and a one-day VaR of 2.24% at the 99% level.

So far, both bonds have been assumed to be independent of each other. However, empirical examinations exhibit strong dependence among the Y_i and the U_i. Therefore, the following section presents the results of the model in 3.7 incorporating dependence between the U_i and dependence between the Y_i.

3.8.5 VaR-Results For The Dependence Assumption

Results of calculating the Gaussian covariances and correlations between the U_i and Y_i of the given bond portfolio are presented in the following tables 3.9 - 3.12:

$\text{cov}(Y_i, Y_j)\ *10^{-4}$	Y_1	Y_2
Y_1	0.7699	0.5821
Y_2	0.5821	0.4785

Table 3.9: cov(Y_i,Y_j), i,j = 1,2

The modeling of the dependent case - as demonstrated in the previous section - is performed by splitting both the Y_i and the U_i into two components. The first

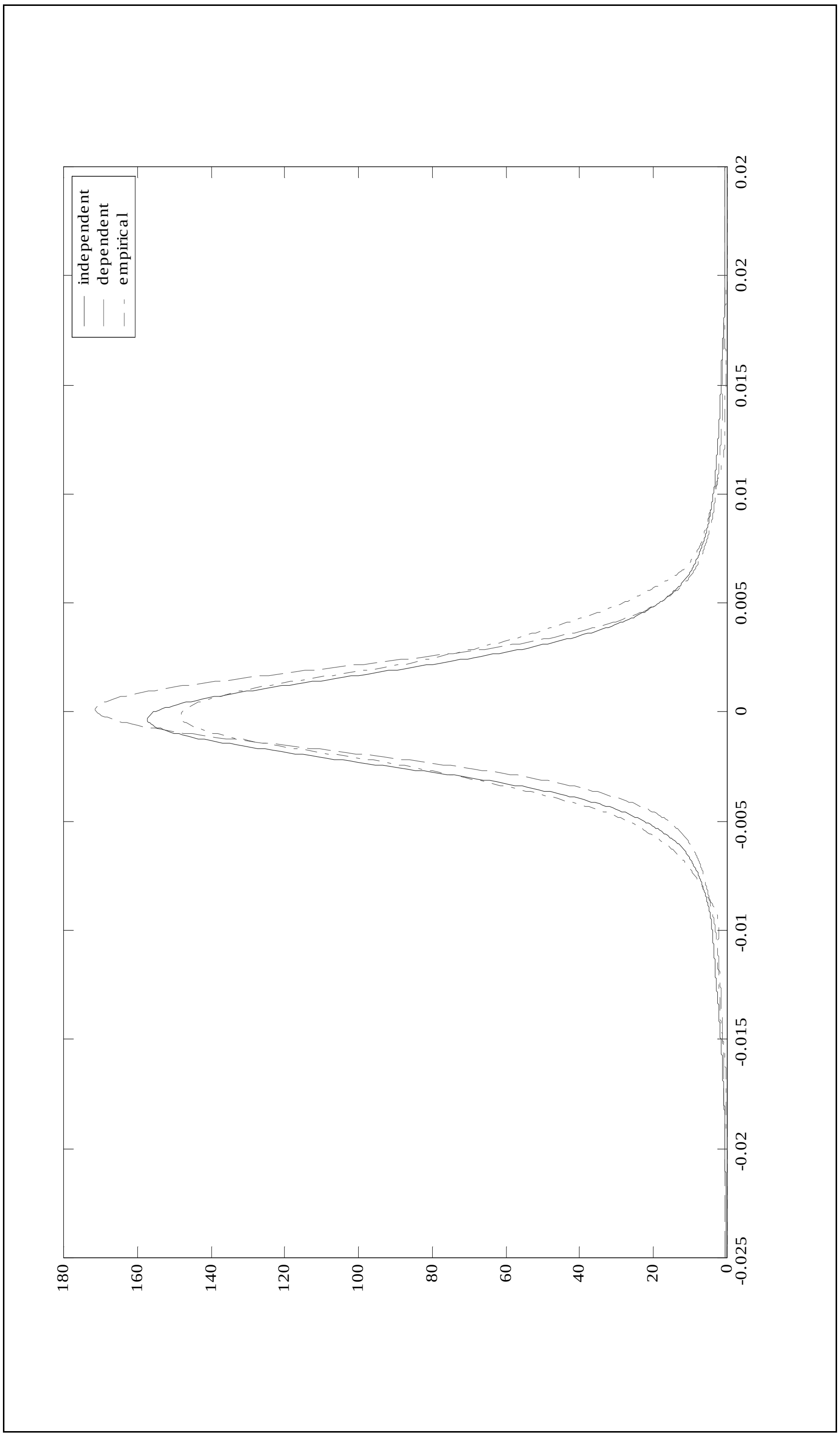

Figure 3.1: Stable models for the independent and dependent cases: density of daily log returns (portfolio).

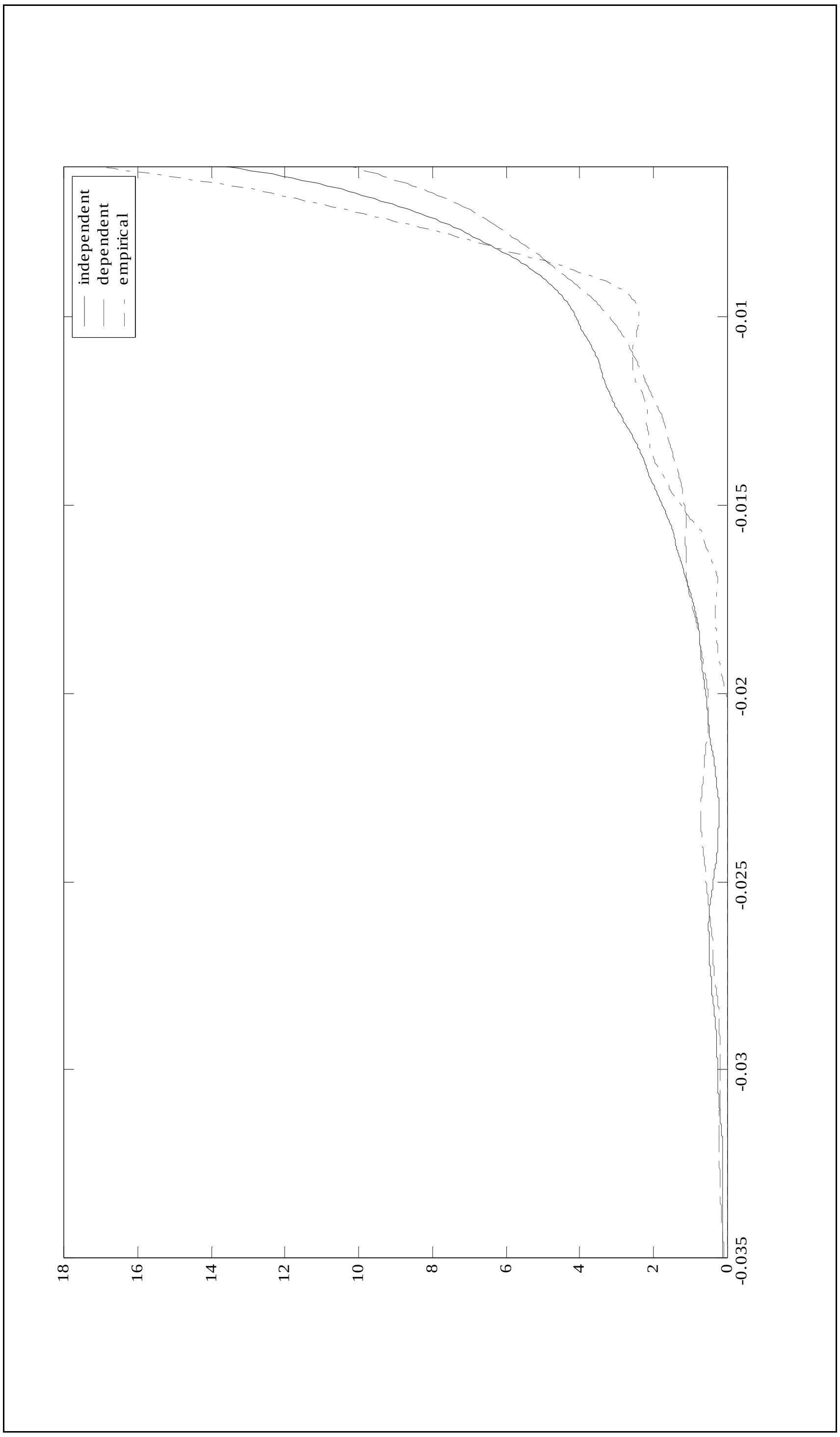

Figure 3.2: Stable models for the independent and dependent cases: density of daily log returns (portfolio) - tails on left-hand side.

$\text{cov}(U_i, U_j) * 10^{-4}$	U_1	U_2
U_1	0.1038	0.0850
U_2	0.0850	0.0748

Table 3.10: cov(U_i,U_j), i,j = 1,2

$\text{cor}(Y_i, Y_j)$	Y_1	Y_2
Y_1	1.0000	0.9591
Y_2	0.9591	1.0000

Table 3.11: cor(Y_i,Y_j), i,j = 1,2

$\text{cor}(U_i, U_j)$	U_1	U_2
U_1	1.0000	0.9653
U_2	0.9653	1.0000

Table 3.12: cor(Y_i,Y_j), i,j = 1,2

component includes the dependence, which is modeled by a sub-Gaussian random vector. The second component illustrates the skewness (see section 3.7).

Table 3.13 provides the Value at Risk (VaR) for the 95% and 99% level with a horizon of one day, comparing both stable models (independent and dependent cases) with the empirical data.

	95%	95%	99%	99%
	log price change	perc. change	log price change	perc. change
Empirical	0.0054	0.54%	0.0108	1.08%
Dependent	0.0060	0.60%	0.0242	2.40%
Independent	0.0068	0.67%	0.0226	2.24%

Table 3.13: Stable portfolio VaR (one day) as log price and percental price changes.

For comparison, table 3.14 presents the VaR, assuming the returns follow a Gaussian law.

	95%	95%	99%	99%
	log price change	perc. change	log price change	perc. change
Dependent	0.0061	0.61%	0.0087	0.87%
Independent	0.0044	0.44%	0.0063	0.63%

Table 3.14: Gaussian portfolio VaR (one day) as log price and percental price changes.

The outcome of the empirical examination can be summarized as follows:

- The results for VaR confirm the earlier findings[18] that for credit returns the Gaussian VaR is only acceptable for the 95% level, but underestimates the 99% level.
- The stable VaR is also appropriate for the 95% level, but it is a more conservative measure for the 99% level. This is actually good because the empirical VaR tends to underestimate the true VaR due to the low number of observations in the tails.[19]
- The Gaussian dependent case achieves much better results than the Gaussian independent case, however it still underestimates the 99% VaR.

Calculating the VaR for longer horizons also, e.g. 10 days, one would necessitate building the 10-day returns for both the corporate bonds and their corresponding treasury bonds from the empirical data and fitting the above models with the data. It is essential to point out that longer horizons cannot be calculated by taking the one-day return model and extending it to the desired horizon simply by applying a Lévy process with independent increments. Subsequent observations of the credit returns are not i.i.d., as *volatility clustering* can be observed and *long-memory effects* might occur. Thus, the volatility for a multiple-day horizon cannot be obtained by a simple scaling approach.[20] Longer forecast horizons require new types of models, while sample data should be available for longer periods.

So far the phenomenon of heavy tails in credit returns has been discussed. Two other aspects, volatility clustering and long-range dependence, have been mentioned now. The following chapter of this work introduces different types of stochastic processes that can be applied for modeling the behavior of financial prices and returns. After introducing *AR(I)MA* and *GARCH* processes, the phenomenon of *long-range dependence (LRD)* is discussed. Both the theory and the methods for detecting and measuring LRD are treated. Finally, credit return data are analyzed for the presence of long-range dependence.

[18] Rachev, Schwartz and Khindanova (2001).
[19] See Khindanova, Rachev and Schwartz (1999).
[20] See Christoffersen, Diebold and Schürmann (1998).

Chapter 4

The Economics of Financial Prices

This chapter begins with a review of basic stochastic processes applied to the modeling of financial prices and then continues with recent advances in this field. First, it provides an explanation of (univariate) ARMA and ARIMA models (the latter in the non-fractional case). Furthermore, it introduces (univariate) GARCH. These processes are possible alternatives for modeling the behavior of financial returns. For both AR(I)MA and GARCH, the stable assumption is considered as well. Finally, a combination of ARMA and GARCH, designated as ARMA with GARCH in errors, is presented as it has become popular in financial applications. A different type of processes are the so-called *subordinated models*. In general, these are a combination of two processes, i.e. a price process and a trading-time process (intrinsic time process) of a financial instrument, where the resulting process is subordinated to the price process by the intrinsic time process.

4.1 Overview

The most famous ideas in forecasting financial prices are the *martingale* and the *random walk*.

The theory of the martingale says that if the history of prices for a financial asset is known up to today, then the expected price for tomorrow is the price of today.[1] Therefore, the expected return is zero. The martingale is a powerful tool but it stands in contradiction to the idea that higher risk requires a higher expected return.

The simplest version of a random walk is the one with independent, identically distributed increments (i.i.d.). For a time series of absolute prices P_t, this random walk model can be represented as

[1]Campbell, Lo and MacKinlay (1997).

$$P_t = \mu + P_{t-1} + \epsilon_t, \tag{4.1}$$

with ϵ_t i.i.d. $N(0, \sigma^2)$. μ is the expected change or drift. (4.1) represents the simplest version of a random walk. Normality is the most common distributional assumption for the ϵ_t. With ϵ_t i.i.d. $N(0, \sigma^2)$, the process is called an *Arithmetic Brownian Motion* sampled at regularly spaced unit intervals. The conditional mean of such a random walk is $E[P_t|P_0] = P_0 + \mu_t$. The variance at time t and the conditional variance is $Var[P_t|P_0] = \sigma^2 t$. From this it is apparent that a random walk is non-stationary.

As is the case in reality, prices P_t cannot become negative: log prices $p_t = log(P_t)$ are usually applied. Otherwise, there would be a positive probability that P_t could become negative. Thus, the random walk becomes

$$p_t = \mu + p_{t-1} + \epsilon_t, \tag{4.2}$$

with ϵ_t i.i.d. $N(0, \sigma^2)$. The assumption of i.i.d. increments does not hold with prices for financial assets. The increments are still assumed to be independent but not identically distributed (they are so-called INID increments). This type of random walk allows unconditioned heteroscedasticity in the ϵ_t. Therefore, it has the ability to model a time-variant volatility. A third version of random walk allows dependence among its increments although the increments remain uncorrelated ($cov[\epsilon_t, \epsilon_{t-k}] = 0$). However, the squared increments are correlated ($cov[\epsilon_t^2, \epsilon_{t-k}^2] \neq 0$).

In the following sections several types of stochastic processes to describe the price behavior of financial assets are discussed. The introduced processes incorporate dependence on past realizations of their variables and disturbances.

Serial dependence in prices of financial assets is very common. Expected future returns are dependent on lagged realizations. This differs from the concepts of random walk and martingale, where the current realization of the variable contains all forecasting information. These forecasting models are termed *conditional models*. The random part is represented by one or a series of stochastic error terms.

ARMA (Autoregressive Moving Average) models are prominent for predicting short-term values of time series. Such models condition the mean on past values. For GARCH (General Autoregressive Conditional Heteroscedasticity) models, the variance of the process is also conditioned by past realizations (heteroscedasticity) of variance and error term. This is different from ARMA models, where the variance is constant over time. The variances of ARMA models are conditionally homoscedastic and do not hinge on past realizations.

AR(I)MA and GARCH models are described in more detail in the following two sections.

4.2 The Application Of ARMA Models For Credit Returns

Autoregressive moving average models ARMA(p,q) have both an autoregressive term of order p, and a moving average term of order q. They are called mixed models. The ARMA process is stationary with the assumption $\mu = E(X_t) = 0$ and represented by the equation

$$X_t = \Phi_1 X_{t-1} + ... + \Phi_p X_{t-p} + \Theta_1 \epsilon_{t-1} + ... + \Theta_q \epsilon_{t-q} + \epsilon_t \tag{4.3}$$

where p and q are integer numbers. The sequence of the ϵ_t is defined as a sequence of i.i.d. Gaussian random variables. The ARMA process can be expressed with the *backward operator*. The backward operator B is defined as follows:

$$BX_t = X_{t-1} \text{ , } B^m X_t = X_{t-m} \text{ etc. , with } m \in N. \tag{4.4}$$

Therefore, $X_t - X_{t-1} = (1-B)X_t$, $X_{t-1} - X_{t-2} = (1-B)X_{t-1} = (1-B)BX_t$. The *forward operator* F is the inverse operator, defined as $F = B^{-1}$ with $FX_t = X_{t+1}$ and $F^m X_t = X_{t+m}$ with $m \in N$.

The autoregressive operator of order p is defined as

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - ... - \Phi_p B^p. \tag{4.5}$$

This yields

$$\Phi(B) X_t = X_t - \sum_{j=1}^{p} \Phi_j X_{t-j}. \tag{4.6}$$

A pure AR(p) process $X_t = \Phi_1 X_{t-1} + ... + \Phi_p X_{t-p} + \epsilon_t$ can be represented with backward operator as

$$\Phi(B) X_t = \epsilon_t. \tag{4.7}$$

Similarly, the MA(q) process can be written in backward operator form:

$$\Theta(B) \epsilon_t = X_t. \tag{4.8}$$

Expressing the complete ARMA(p,q) process by using backward operators only, one obtains

$$\Phi(B) X_t = \Theta(B) \epsilon_t. \tag{4.9}$$

Usually, time series assumed to follow a Gaussian ARMA process are investigated using the Box-Jenkins method.[2] The Box-Jenkins method analyzes the underlying stochastic process of a time series. The data process is considered to be time invariant, and its parameters are not supposed to change. The Box-Jenkins method tries to detect the inner relationship of successive observations in order to construct an optimal forecast function. The types of processes the Box-Jenkins method deals with are (i) stationary processes or (ii) homogeneous instationary processes.

Box-Jenkins is an iterative procedure that is performed in five stages:

- Explorative data analysis
- Model identification
- Model estimation
- Model evaluation
- Prognosis

A standard method for model identification is the calculation of the time series' sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF). The obtained values are then compared with the theoretical autocorrelation function (ACF) and the theoretical partial autocorrelation function (PACF) of alternative models. The theoretical ACF with lag k for a time series y_t is defined as [3]

$$\gamma_k = \frac{E[(y_t - \mu_y)][(y_{t+k} - \mu_y)]}{\sqrt{E[(y_t - \mu_y)^2]E[(y_{t+k} - \mu_y)^2]}} = \frac{cov(y_t, y_{t+k})}{\sigma_{y_t}\sigma_{y_{t+k}}}, \tag{4.10}$$

with μ_y as the theoretical mean and σ_{y_t} the standard deviation of the time series. For a stationary process, the variance at time t in the denominator of equation (4.10) is the same as the variance at time $t + k$. Therefore, the denominator becomes the variance of the process y_t.

$$\gamma_k = \frac{cov(y_t, y_{t+k})}{\sigma_y^2} \tag{4.11}$$

γ_k in equation (4.10) is the theoretical ACF. In practice, there is only a limited number of observations available. For this reason, an estimate of the theoretical ACF, called the sample autocorrelation function (SACF), has to be calculated:

[2] Box and Jenkins (1976).

[3] Pindyck and Rubinfeld (1991, p. 446).

$$\hat{\gamma}_k = \frac{E[(y_t - \bar{y})][y_{t+k} - \bar{y}]}{\sqrt{E[(y_t - \bar{y})^2]E[(y_{t+k} - \bar{y})^2]}}, \tag{4.12}$$

with $\bar{y}$ as the sample mean of the available observations of the time series.

Some information about the order of an autoregressive (AR) process can be obtained from the oscillatory behavior of the SACF; much more information can be obtained from the partial autocorrelation function (PACF).

The m-th partial autocorrelation is defined as the last coefficient $\alpha_m^{(m)}$ in a linear projection of y on its m most recent values.[4] An estimate of the m-th partial autocorrelation is the last coefficient of the following OLS regression:

$$y_{t+1} = \hat{c} + \hat{\alpha}_1^m y_t + \hat{\alpha}_2^m y_{t-1} + ... + \hat{\alpha}_m^m y_{t-m+1} \ . \tag{4.13}$$

In case of an AR(p) process, $\hat{\alpha}_m^m$ would be zero for $m = p + 1$, $m = p + 2$, ... However, for an MA(q) process $\hat{\alpha}_m^m$ would asymptotically approach zero with no abrupt cut-off.

Assume an AR(1) process. Then the ACF shows an exponential decay, whereas the PACF only differs from zero at $\hat{\alpha}_1^m$. The PACF provides information on the order of the autoregressive part of the process.

Because of the duality between MA(1) and AR(1) processes, for the MA(1) the behavior of ACF and PACF is the other way round compared to AR(1). Following the Box-Jenkins method, a model fit is performed in three stages:

- Order selection (by applying the SACF, SPACF or the Akaike Information Criterion).
- Estimation of the parameters.
- Diagnostic checking.

The basic idea of Box-Jenkins is Wold's theorem: any stationary time series generating process can be represented by an infinite dimensional MA process.

Combining MA and AR representations usually allows a lower number of parameters to be used.

The Box-Jenkins method can also be applied when the innovations are assumed to follow a stable law with $\alpha < 2$. However, for $\alpha < 2$, the ACF works poorly, as the rate of distributional convergence for the autocorrelations is very slow (compared to the case $\alpha = 2$).

For data exhibiting sudden bursts of amplitude, the Gaussian ARMA process is not appropriate. There are two alternatives to encapsulate the burstiness with an ARMA model:[5]

[4] See Hamilton (1994).
[5] Gallagher (1998).

1. One gives up the linear structure of the process and moves to a nonlinear time series.

2. The other replaces the Gaussian innovations by a more general class of distributions, stable distributions, which show heavy-tailedness and peakedness for $\alpha < 2$.

For an extensive coverage of the Box-Jenkins method in order to specify stable ARMA, see Gallagher's dissertation work (1998). Additionally, the reader is referred to Samorodnitsky and Taqqu (1994, pp. 376) for more on stable ARMA.

4.3 Nonfractional ARIMA Models

An ARIMA (p,d,q) process, called autoregressive *integrated* moving average process, is an ARMA(p,q) process with the variable X differenced d times. ARIMA(p,d,q) processes are a generalization of ARMA(p,q) processes. As empirical time series may exhibit instationary variables with trends in their mean and variance, the non-stationarity can be removed by differencing the time series in order to render it stationary. This is necessary because a process generated by an ARMA(p,q) model is always stationary.

d is an integer variable and describes the number of differencings applied to X_t:

$$w_t = \Delta^d X_t \ , \tag{4.14}$$

where X_t is a homogeneous non-stationary process of order d. Δ is the *difference operator* and d is the *order of differencing*. For example:

$$\Delta X_t = X_t - X_{t-1} = (1-B)X_t \text{ where d} = 1; \tag{4.15}$$

$$\Delta^2 X_t = \Delta X_t - \Delta X_{t-1} = (1-B)X_t - (1-B)X_{t-1} = (1-B)^2 X_t \text{ where d} = 2; \tag{4.16}$$

This leads to the following representation of the ARIMA(p,d,q) process (in backward operator form):

$$\Phi(B)\Delta^d X_t = \Theta(B)\epsilon_t, \text{ or} \tag{4.17}$$

$$\Phi(B)X_t = \Theta(B)\Delta^{-d}\epsilon_t, \tag{4.18}$$

where the sequence of innovations ϵ_t is defined as a sequence of i.i.d. Gaussian random variables. For non-fractional ARIMA processes, d, which denotes the number of differencings, can only take integer values.

ARIMA(p,d,q) models can be applied to time series with a long-term growth trend component. It depends on the time series whether higher than first-order differences are required to obtain stationarity (see equation (4.14).

The stationary time series w_t yields an ARMA(p,q) process. X_t is driven by a classical (non-fractional) ARIMA(p,d,q) with an integer-differencing. ARIMA models can be further generalized by allowing fractional noise, which is the case for non-integer d. Such *fractional ARIMA* models are introduced later in this work.

When dealing with financial time series, the notion of homoscedasticity is rarely true.[6] For empirical return data, researchers often found that large (absolute) returns are often followed by large (absolute) returns. The variance of the process is usually conditioned by past realizations. These processes show signs of clustered volatility.

4.4 Modeling Credit Risk With GARCH(p,q)

Observations of the credit returns R_i and of the risk spread U_i exhibit clustered volatility.[7] This indicates heteroscedasticity so that the application of *autoregressive conditional heteroscedastic (ARCH)* models seems to be promising. The basic autoregressive conditional heteroscedastic model ARCH(q) was developed by Engle in 1982.[8] For a time series X_t with mean μ, it is described by the following equations:

$$\begin{aligned} X_t - \mu &= \sigma_t \epsilon_t \text{ , with} \\ \sigma_t^2 &= f_0 + \sum_{i=1}^{q} f_i \epsilon_{t-i}^2 \text{ ,} \end{aligned} \tag{4.19}$$

where q is a positive integer and $\epsilon_t \sim N(0,1)$. The size of the error term's variance σ_t^2 is determined by the q previous representations of the squared error term ϵ_t^2.

GARCH(p,q) is a general form of of ARCH(q), where the variance parameter is also determined by the p previous representations of the conditional variance σ_t^2:

[6] Mittnik, Rachev and Paolella (1997).
[7] Rachev, Schwartz and Khindanova (2001).
[8] Bollerslev (1992); Engle (1982).

$$X_t = \sigma_t \epsilon_t, \text{ with}$$
$$\sigma_t^2 = f_0 + \sum_{i=1}^{q} f_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} g_j \sigma_{t-j}^2 \ , \tag{4.20}$$

where $\epsilon_t \sim N(0,1)$. Compared to linear-dependent AR and MA processes, (G)ARCH models are nonlinear. Moreover, their frequency distribution is peaked and heavy-tailed, even if their innovations are Gaussian.

(G)ARCH models show a time-dependent variance. In case of ARCH, the expected variance of the model is conditioned by previous representations of the model's squared error terms. In case of GARCH, it is conditioned by previous representations of the squared error term and previous representations of the variance itself. Conditional heteroscedasticity is given by the proceeding time.

One of the most striking properties of (G)ARCH is the fact that small changes will most likely be followed by small changes. Hence, the signs of the changes are unpredictable. The plot of such a process exhibits different volatility clusters.

Because of the nonlinearity, large changes amplify and small changes contract. This results in fat tails and peaked distributions for (G)ARCH models.

Focusing on the model for credit returns $R_{i,t} = a_{i,t} + b_{i,t}Y_{i,t} + U_{i,t}$ (introduced in section 3.2), the empirical plot of the $\hat{U}_i$ exhibits fat tails and peakedness.[9] Moreover, a clustering behavior of their volatilities can be observed.

Applying a GARCH representation for the $\hat{U}_i$, one obtains:

$$U_{i,t} = \sigma_{i,t}\epsilon_{i,t}, \text{ with}$$
$$\epsilon_{i,t} \sim N(0,1);$$
$$\sigma_{i,t}^2 = f_{0,i} + \sum_{j=1}^{p} g_{i,j}\sigma_{i,t-j}^2 + \sum_{k=1}^{q} f_{i,j}U_{i,t-k}^2. \tag{4.21}$$

GARCH is not able to meet properties such as self-similarity (see equation (5.1)). It does not comply with fractal processes (see chapter 5), and is rather appropriate when investigating period-specific data.

A stationary GARCH process X_t is defined by $E(X_t) = 0$ and $Var(X_t) = 1$. It can easily be fitted with the log returns of speculative prices. The clustering of the volatility is called dependence in the tails. The Pareto-like marginal distribution shows a power-law behavior: $P(X > x) \sim c_0 x^{-k}$, as $x \to \infty$, for some $c_0, k > 0$.

A GARCH process might turn out to show tiny autocorrelations among its data.[10] But even more, the sample autocorrelations of absolute and squared values

[9] Rachev, Schwartz and Khindanova (2001).
[10] Mikosch and Starica (2000a).

are significantly different from zero, and this is the case even for large lags. The ACFs for GARCH-driven log prices decay to zero at exponential rate. Although the sample ACFs of absolute and squared values do not seem to be zero for large lags, which usually indicates a long memory, these values for such ACFs are insignificant due to wide confidence intervals. Therefore, the long-range dependence behavior of the GARCH-process's volatility (represented by the ACFs of the squared values) would not be in contradiction to the short-term memory property of the GARCH process.

Mikosch and Starica (2000a) try to explain this behavior of the sample ACFs of absolute and squared log returns with shifts in the unconditional variance of the model. These shifts are due to changing parameter values of the GARCH model. Therefore, they propose the applied GARCH process should focus on shorter time series, as a change of its parameter values could occur with longer time series.

4.5 Stable GARCH Models

Studies of GARCH-filtered time series often showed that the innovations remain heavy-tailed.[11] A GARCH process with stable innovations is quite common.

Therefore, GARCH and α-stable assumption are not considered to be competing hypotheses but could be complementary ones.

Testing for GARCH. A common way to test for a GARCH behavior of the volatility is calculating the autocorrelation functions of the squared series, i.e. both the sample autocorrelation function (SACF) and the partial autocorrelation function (PACF). The partial autocorrelation helps to determine the order of the autoregressive process. Assume an AR(p) process has an autocorrelation function which is infinite in extent. It can be described by p non-zero functions of the autocorrelations. The PACF of a p-th order process has a cut-off after lag p.

Definition of stable GARCH. A time series X_t , $t \in Z$, follows a stable GARCH (α, p, q) process [12] if

- $X_t = \sigma_t S_t$ with S_t i.i.d. random variables following an $S\alpha S$ distribution with scale 1, $S\alpha S(1)$. $1 \leq \alpha < 2$.
- $\sigma_t = \sigma + \sum_{i=1}^{q} \alpha_i |X_{t-i}| + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$, $t \in Z$. The α_i, with $i = 1...q$, and β_j, with $j = 1...p$, are non-negative constants and $\sigma > 0$.

[11] Rachev and Mittnik (2000, p. 4).
[12] Rachev and Mittnik (2000, p. 282).

As most financial time series show a finite mean, the assumption of $1 < \alpha \leq 2$ is not really restrictive.

Furthermore, the stable GARCH(α, p, q) can be generalized to the *asymmetric* case with the sequence X_t, $t \in Z$, having a time-varying mean.

Definition of asymmetric stable GARCH. An $S_{\alpha,\beta}$ GARCH process is defined if the following conditions are true:

- $X_t = \mu_t + c_t \epsilon_t$ with $\epsilon_t \sim$ i.i.d. $S_{\alpha,\beta}$. The mean of the series is time-varying: $E(X_t) = \mu_t$, which allows for a broad range of mean equations.
- $c_t = a_0 + \sum_{i=1}^{q} \alpha_i |X_{t-i} - \mu_{t-i}| + \sum_{j=1}^{p} \beta_j c_{t-j}$.

$S_{\alpha,\beta}$ denotes the standard asymmetric stable Paretian distribution with stability-index α, skewness parameter $\beta \in [-1, 1]$, zero location parameter and unit scale parameter.

For examples and simulations on stable GARCH processes, the reader is referred to Rachev and Mittnik (2000, chapt. 6), and Mittnik, Rachev and Paolella (1997).

4.6 ARMA Models With GARCH In Errors

ARMA processes are the traditional tools for modeling of serial dependence, and GARCH processes stand for volatility clustering. Both effects can be combined in so-called ARMA-GARCH models, an ARMA process whose innovations are driven by GARCH processes. The model is represented by the following equations:

$$X_t = \mu + \sum_{i=1}^{p} a_i X_{t-1} + \epsilon_t + \sum_{j=1}^{q} b_j \epsilon_{t-j} \ , \tag{4.22}$$

$$\epsilon_t = c_t u_t \text{ , with} u_t \text{ i.i.d.}$$

For Gaussian innovations $u_t \sim N(0, 1)$:

$$c_t^2 = w + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j c_{t-j}^2$$

For stable innovations $u_t \sim S_\alpha(1)$, this yields

$$c_t = w + \sum_{i=1}^{r} \alpha_i |\epsilon_{t-i}| + \sum_{j=1}^{s} \beta_j c_{t-j} \ ,$$

with $c_t > 0$.

For $q = 0$ this becomes an AR-GARCH model. ARMA-GARCH or AR-GARCH models are popular for fitting financial returns as they absorb both heteroscedasticity and serial correlation. The identification procedure is identical to that for separate ARMA and GARCH models.

For stable ARMA-GARCH, the conditional distributions usually show a greater α than for the unconditional distributions because the ARMA-GARCH components partially absorb the kurtosis.[13]

4.7 Subordinated Models

Another way to model the dynamics of an asset price or index is to apply the recognized lognormal price model, which is driven by Brownian Motion, and add one more constant parameter to it. The physical time in the stochastic process that governs the lognormal asset price model is substituted by an *intrinsic time*. The intrinsic time, which represents the *trading time* (i.e. the number of trades effected up to a certain point of time), has the ability to provide tail effects as observed in the market.

The method of subordination works with a stochastic process $W = \{W(T), T \geq 0\}$, for the behavior of the log price and with a non-decreasing stochastic process $T = \{T(t), t \geq 0\}$ describing the trading time.[14] The resulting process is $W(T(t))$, $t \geq 0$.

The process of $W(T(t))$ is said to be *subordinated* to W by the intrinsic time process $T(t)$.

The price S(t) of an asset can be modeled by a stochastic process of the following form:

$$log(S(t)) = log(S(t_0)) + \mu(t - t_0) + \rho(T(t) - T(t_0)) + \sigma(W(T(t)) - W(T(t_0))). \quad (4.23)$$

The noise process is described by a standard Wiener process W. W and T are assumed to be independent of each other. The intrinsic time denoted by T represents the market activity. Higher market activity results in higher price changes. Market time evolves at a different rate, depending on the information flow occurring.

The resulting stochastic process $Z = W(T(t))$, $t \geq 0$, is the new driving process for the lognormal model. μ represents the drift in physical time, ρ the drift in intrinsic time, and σ the volatility.

[13] Mittnik, Rachev and Paolella (1997).

[14] See Marinelli, Rachev and Roll (2001) and Marinelli, Rachev, Roll and Göppl (1999).

The intrinsic time process $T(t)$ can be designated as the stochastic deformation of the time scale. It is assumed that the price process $W(t)$ is a Standard Brownian Motion. The skewness of the log price distribution is caused by the intrinsic time process.

Assume the scaling behavior is determined by a maximally skewed ($\beta = 1$) $\alpha/2$-stable process with $\alpha \in]0,2[$ having independent increments:

$$T(t) - T(s) \sim S_{\alpha/2}(c|t-s|^{\alpha/2}, 1, 0),$$
$$c = 2cos\left(\frac{\pi\alpha}{4}\right)^{2/\alpha}. \tag{4.24}$$

c is chosen so that increments are closed under convolution. For $Z(t) = W \circ T(t) = W(T(t))$, $T \in R_0^+$, one obtains a standard symmetric α-stable Lévy-process with stationary independent increments, following the symmetric α-stable distribution $S_\alpha(|t-s|^{\frac{1}{\alpha}}, 0, 0)$. Although the process W is Gaussian, $Z(t)$ exhibits heavy-tailed increments.

Return to the log price model yield

$$log(S(t)) = log(S(t_0)) + \mu(t-t_0) + \rho(T(t)-T(t_0)) + \sigma(W(T(t)) - W(T(t_0))). \tag{4.25}$$

The differences in the logarithmic price between physical time t and s, which are the increments of the log price process, follow a symmetric α-stable distribution

$$S_\alpha(\sigma|t-s|^{1/\alpha}, 0, \mu|t-s|). \tag{4.26}$$

This model is referred to as the *log-stable* model. For α moving towards 2, it provides the classical lognormal model with the scaling behavior $\sigma|t-s|^{1/2}$ for the distribution of log price differences.

For more on subordination and the log-stable model, see Hurst, Platen and Rachev (1997).

Chapter 5

Long-Range Dependence In Financial Time Series

The first part of this chapter introduces the theory of fractionally integrated processes and *long-range dependence (LRD)*. The second part explains methods for detection and measurement of LRD within a time series. Finally, the tests are applied to time series of the returns of corporate bond indices. In this empirical example, half of the observed bond indices are found to exhibit significant LRD.

All the stochastic processes discussed in the previous chapter are characterized by an integer order of integration. The processes considered now have a non-integer order of integration, with some of them incorporating a phenomenon called *long-range dependence* or *long memory*.

Time series can have a long memory. These systems are not i.i.d. This phenomenon is often referred as *burstiness* in the literature.[1] The underlying stochastic processes for such burstiness are called *fractal* . Fractal processes with a long memory are denoted as persistent. A common characteristic of such fractal processes is that their space time is governed parsimoniously by power law distributions. This effect is called the *Noah Effect* and explains the occurrence of heavy tails and infinite variance. It can be observed as the tendency of time series toward abrupt and discontinuous changes. Another property of fractal processes are their hyperbolically decaying autocorrelations, which is known as the *Joseph Effect*. This is the tendency of a persistent time series to have trends and cycles.

For a long-memory process, larger-than-average representations are more likely to be followed by larger-than-average representations instead of lower-than-average representations. Hurst developed a statistic to examine the long memory of a stochastic process. As significant autocorrelations are frequently not visible, he came up with a new methodology to provide a measure, the so-called *Hurst Expo-*

[1] Willinger, Taqqu and Erramilli (1996).

nent, for long-range dependence within a time series.[2]

The examination of fractal processes in finance has become a popular topic over the years.[3]

5.1 Self-Similar Processes

Due to the failures of traditional capital market theory, which is largely based on the theory of martingales, researchers experienced that markets do not follow a purely random walk. Hence, the fractal market hypothesis, where the existence of *self-similar structures* comprises a major component, was developed. For self-similar processes, small increments of time are statistically similar to larger increments of time.

Self-similarity is defined as follows:[4] Let X_t be a stochastic process with a continuous time t. X_t is self-similar with self-similarity parameter H *(H-ss)*, if the rescaled process with time scale ct, $c^{-H}X_{ct}$, is equal in distribution to the original process X_t.

$$X_t \stackrel{\mathrm{d}}{=} c^{-H} X_{ct}. \tag{5.1}$$

Hence, for a sequence of time points $t_1, ..., t_k$ and a positive stretch factor c, the distribution of $c^{-H}(X_{ct_1}, ...X_{ct_k})$ is identical to that of $X_{t_1}, ...X_{t_k}$. In other words, the path covered by a self-similar process always looks the same, regardless of the scale it is observed with. In terms of financial data this means that no matter whether the data in question was intraday, daily, weekly, or monthly data, the plots of the resulting processes have similar looks.

For a self-similar process, its limiting behavior (t moving towards infinity) exhibits the following properties, depending on the value of H.

$$X_t \to 0 \text{ if } H < 0, \tag{5.2}$$

$$X_t \to X_1 \text{ if } H = 0, \tag{5.3}$$

$$X_t \to \infty \text{ if } H > 0, \tag{5.4}$$

where $\to$ means convergence in distribution. A non-degenerate H-ss process cannot be stationary. As a stationary process would show equality in distribution at any

[2] See Hurst (1951).

[3] For example, see Mandelbrot (1997a), Mandelbrot (1997b), Mandelbrot (1999) and Peters (1994).

[4] Beran (1994).

point of time, one obtains $X(t) \stackrel{\mathrm{d}}{=} X(at)$, $t > 0, a > 0$. On the other hand, for an H-ss process $X(at) \stackrel{\mathrm{d}}{=} a^H X(t)$ holds true. However, for $a \to \infty$ this yields $a^H X(t) \to \infty$ Therefore, a non-degenerate H-ss process cannot be stationary.[5]

There is a key correspondence between self-similarity and stationarity:

- If $X(t)$, $0 < t < \infty$, is an H-ss process, then $Y(t) = e^{-tH} X(e^t)$, $-\infty < t < \infty$ is stationary.
- Conversely, if $Y(t)$, $-\infty < t < \infty$ is stationary, then $X(t) = t^H Y(log(t))$, $0 < t < \infty$, is H-ss.

For example, the $S\alpha S$-Lévy-motion X, $0 < \alpha < 2$ is H-ss with self-similarity parameter $H = 1/\alpha \in (0.5, \infty)$. $X = \{X(t), t \in T\}$ has independent increments:

$$X(t) - X(s) \sim S_\alpha \left(|t-s|^{1/\alpha} \right). \tag{5.5}$$

The concept of self-similarity is of interest in the presence of *scale invariance*. The idea of scale-invariance in financial data goes back to Mandelbrot, who maintains that, for a chosen time scale T, the distribution of the price changes can be calculated from a distribution for a shorter time scale $\tau < T$.

$$P_T(x) = \frac{1}{\lambda} P_\tau \left(\frac{x}{\lambda} \right) \text{ with } \lambda = \left(\frac{T}{\tau} \right)^H, \tag{5.6}$$

where H is the self-similarity parameter.

For further information on self-similarity the reader is referred to Samorodnitsky and Taqqu (1994, chapt. 7), or Beran (1994).

In the following section common fractional processes are explained: *Fractional Brownian Motion* and a fractional process with stable innovations, *Fractional Lévy Motion*.

5.2 Fractional Processes And The Hurst Exponent

First, consider a process without a long memory. A perfect example is *Standard Brownian Motion*, which is characterized as a standard random walk.[6] Commonly known is Einstein's *to the one-half* rule, which describes the distance covered by a particle driven by Standard Brownian Motion. It states that the distance between

[5] Samorodnitsky and Taqqu (1994, pp. 312).

[6] See Campbell, Lo and MacKinlay (1997).

consecutive values of the observed time series of this particle is proportional to the square root of time:[7]

$$R \sim T^{0.5}. \tag{5.7}$$

The power of 0.5 refers to the Hurst Exponent which is also known as the self-similarity parameter. For Standard Brownian Motion, the Hurst Exponent H is equal to 0.5, which means that it is an unbiased random walk. A process with a Gaussian limiting distribution but a Hurst Exponent H different from 0.5 is called Fractional Brownian Motion. Fractional Brownian Motion differs from Standard Brownian Motion in that it is a biased random walk. The odds are biased in one direction or the other.

5.2.1 Stationary Increments

Definition. A process with real values $X(t)$, $0 < t < \infty$ has stationary increments if

$$\{X(t+h) - X(h), t \in T\} \overset{\mathrm{d}}{=} \{X(t) - X(0), t \in T\}, \text{ for all } h \in T. \tag{5.8}$$

A process $\{X(t),\ t \in T\}$, is called H-sssi if it is *self-similar with index H and has stationary increments*. Brownian Motion is H-sssi with $H = 1/2$ and α-Lévy Motion is H-sssi with $H = 1/\alpha$.

Fractional Brownian Motion is the unique H-sssi process for the Gaussian case, which is important in the context of long-range dependence. The properties of Fractional Brownian Motion are described in the following.

5.2.2 Definition of Fractional Brownian Motion

Assume a self-similar Gaussian process with $X_t, t \in R$ having mean zero and the autocovariance function

$$Cov(X_{t_1}, X_{t_2}) = \frac{1}{2}(|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H})VarX(1), \tag{5.9}$$

where H is the self-similarity parameter and $H \in (0, 1)$.

Such a process is called a Fractional Brownian Motion. The process is H-sssi. For $H = 0.5$ it becomes a Standard Brownian Motion.

The increments of Fractional Brownian Motion, $Y_j = B_H(j+1) - B_H(j)$, $j \in Z$, form a stationary sequence Y_j, which is called *Fractional Gaussian Noise*.[8]

[7] Peters (1994, p. 55).

[8] See Samorodnitsky and Taqqu (1994, pp. 318).

5.2.3 Definition of Fractional Gaussian Noise

A sequence of Fractional Gaussian Noise has the following properties:

- (i) its mean is zero,
- (ii) its variance $EY_j^2 = EB_H^2(1) = \sigma_0^2$, and
- (iii) its autocovariance function is $r(j) = \frac{\sigma_0^2}{2}[(j+1)^{2H} - 2j^{2H} + (j-1)^{2H}]$, where $j \in Z$, $j \geq 0$, and $r(j) = r(-j)$ for $j < 0$.

For $j \to \infty$, $r(j)$ behaves like a power function.

$$\lim_{j \to \infty} r(j) \to 0. \tag{5.10}$$

The autocorrelations are given by

$$\rho(j) = \frac{1}{2}[(j+1)^{2H} - 2j^{2H} + (j-1)^{2H}], \tag{5.11}$$

where $j \geq 0$ and $\rho(j) = \rho(-j)$ for $j < 0$. As j tends to infinity, $\rho(j)$ is equivalent to $H(2H-1)j^{2H-2}$.

In the presence of long memory, $0.5 < H < 1$, the correlations decay to zero so slowly that they are no longer summable:

$$\sum_{j=-\infty}^{\infty} \rho(j) = \infty. \tag{5.12}$$

For $H = 0.5$, i.e. a Gaussian i.i.d. process, all correlations at non-zero lags are zero. For $0 < H < 0.5$, the correlations are summable, and it holds that

$$\sum_{j=-\infty}^{\infty} \rho(j) = 0. \tag{5.13}$$

$H = 1$ implies $\rho(j) = 1$. For $H > 1$, the condition $-1 \leq \rho(j) \leq 1$ is violated.

For $0 < H < 1$, a Gaussian process with mean zero and the given autocovariance function is self-similar and has stationary increments ($H - sssi$). The above autocovariance function is shared by all Gaussian H-sssi processes.

5.2.4 Fractional Processes With Stable Innovations

There are many different extensions of the Fractional Brownian Motion to the α-stable case with $\alpha < 2$. Most common is the so-called Linear Fractional Stable Motion or, Linear Fractional Lévy Motion.

In an analogy to the Gaussian case with $\alpha = 2$, the increments of Linear Fractional Stable Motion[9] show long-range dependence for $H > 1/\alpha$. LRD for $\alpha < 1$ does not exist, as H must lie in (0,1). Processes with $H = 1/\alpha$ are called α-stable Lévy Motion whose increments $X(t_{j+1}) - X(t_j)$ are all mutually independent.

For the stationarity of the increments of a Lévy Motion it is sufficient to show $X(at) - X(as) \stackrel{d}{=} a^{1/\alpha}(X(t) - X(s))$.

For α-stable Lévy processes with infinite variance, the value obtained for H has to be interpreted carefully due to its relation to the parameter d indicating the degree of long-range dependence.

H, the Hurst Exponent, is the scaling parameter and describes asymptotical self-similarity:

For finite variance processes, the relation between H and d is

$$H = d + 1/2. \tag{5.14}$$

For processes with infinite variance ($\alpha < 2$), the relation is

$$H = d + 1/\alpha. \tag{5.15}$$

If $d > 0$, the time series is governed by a long-memory process.

Cont, Potters and Bouchaud (1992) use the S&P index futures to analyze the scaling behavior of the price increments of high-frequency financial data. They observe anomalous scaling properties for the kurtosis, indicating departure from the i.i.d. hypothesis. Although the ACF of the price changes decays rapidly, the squares exhibit a power law. The square of the increments shows long-memory properties.

Mandelbrot proposed that the increments follow an α-Lévy-Motion with self-similarity index $H = 1/\alpha$. But for longer time series, the properties of scale invariance disappear and therefore the scaling properties are no longer i.i.d. Cont, Potters, and Buchoud propose truncated Lévy flight models.

If the ACF of price changes decays quickly with increasing lag number, the absence of significant linear correlations in asset returns is usually a support for the Efficient Market Hypothesis (EMH). The fast decay of the ACF implies the additivity of the variances (in the Gaussian case).

[9] Samorodnitsky and Taqqu (1994, p. 343).

This section has provided some theoretic background about self-similarity and fractional processes. The following section will introduce methods to detect and measure long-range dependence (LRD) in financial data.

5.3 Detecting and Measuring LRD

There are a number of methods to distinguish a purely random time series from a fractional one. Some directly measure the Hurst Exponent H, others perform a hypothesis test to check the presence of long memory. For example, the *classical R/S analysis* determines the parameter H of a time series.[10] The resulting graph is called *pox-plot of R/S* or *rescaled adjusted range plot*. Other methods to determine the Hurst Exponent are the *Aggregated Variance* method and a similar method called *Absolute Values of Aggregated Series*.[11] Lo's test (1991) is a modification of the classical R/S and checks the null hypothesis *no LRD*. While Lo's test works under the Gaussian assumption, the statistic proposed by Mansfield, Rachev and Samorodnitsky (1999) is independent of the actual tail index α and is thus applicable to both Gaussian and stable time series.

5.3.1 The Aggregated Variance Method

The original time series $X = (X_i, i = 1...N)$ is divided into blocks. Each block is m elements in size. The index k labels the block. The aggregated series is calculated as the mean of each block:

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i \quad \text{with } k = 1, 2, ..., \left[\frac{N}{m}\right]. \tag{5.16}$$

After building the aggregated series, the sample variance of $X^{(m)}(k)$ is obtained as

$$\hat{Var}X^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} (X^{(m)}(k))^2 - \left(\frac{1}{N/m} \sum_{k=1}^{N/m} X^{(m)}(k) \right)^2. \tag{5.17}$$

The procedure is repeated for different values of m $\{m_i, i \geq 1\}$. The chosen values for m should be equidistant on a log scale, i.e. $m_{i+1}/m_i = C$.

As $X^{(m)}$ scales with $m^{(H-1)}$, the sample variance $\hat{Var}X^{(m)}$ behaves identical to $m^{(2H-2)}$. Thus, when plotting a log-log representation of m and $\hat{Var}X^{(m)}$, the

[10] Mandelbrot and Wallis (1968).

[11] Teverovsky, Taqqu and Willinger (1995), and Teverovsky, Taqqu and Willinger (1998).

plots form a straight line with slope $2H - 2$.

5.3.2 Absolute Values Of The Aggregated Series

This method is similar to the Method of Aggregated Variance explained above. Starting again with the aggregated series, the sum of the absolute values of the aggregated series is calculated.

$$\frac{1}{(N/m)} \sum_{k=1}^{N/m} \left| X^{(m)}(k) \right| \ . \tag{5.18}$$

If the original series has a long-range dependence parameter H, the log-log plot of m versus the corresponding values of the statistic yields a line with slope $H-1$.

5.3.3 Classical R/S Analysis

Assume there is a time series of N consecutive values. $Y(n) = \sum_{i=1}^{n} X_i \ ,n \geq 1$, is the partial sum and $S^2(n) = \frac{1}{n}\sum_{i=1}^{n}[X_i - n^{-1}Y(n)]^2$, $n \geq 1$, is the corresponding sample variance.

Define $Z(t) = Y(t) - \frac{t}{n}Y(n)$. The *rescaled-adjusted-range* statistic or R/S statistic is given by

$$\frac{R}{S}(n) = \frac{1}{S(n)}[\max_{0\leq t\leq n} Z(t) - \min_{0\leq t\leq n} Z(t)]. \tag{5.19}$$

R/S is called the rescaled adjusted range as its mean is zero, and it is expressed in terms of the local standard deviation. For large n, the expected value of the statistic approaches $c_1 n^H$:

$$E[R/S(n)] \sim c_1 n^H, \tag{5.20}$$

where c_1 is a positive, finite constant and does not depend on n. In case of long-range-dependence in a Gaussian process, the values for H range in the interval $(0.5, 1.0)$. For an i.i.d. Gaussian process (i.e. pure random walk) or a short-range dependent process, the value of $R/S(n)$ approaches $c_2 n^{0.5}$. c_2 is independent of n, finite, and positive.

$$E(R/S(n)) \sim c_2 n^{0.5}. \tag{5.21}$$

The practical application of the R/S analysis is performed graphically. It is set out in Mandelbrot and Wallis (1968) in detail.

With this procedure, K different estimates of $(R/S(n))$ are obtained by dividing the total sample of N consecutive values into K blocks, each of size N/K.

$$k_{(m)} = \frac{(m-1)N}{K} + 1 \tag{5.22}$$

defines the starting points of each block, where K is the total number of blocks and $m = 1...K$ is the current block number. Now the $R(n, k_{(m)})/S(n, k_{(m)})$ are computed, for each lag n such that $k_{(m)} + n < N$. All data points before $k_{(m)}$ are ignored in order to avoid the influence of particular short-range dependence in the data.

Plotting the $log(R(n, k_{(m)})/S(n, k_{(m)}))$ for each block versus $log(n)$ allows the slope of the fitted straight line to be estimated. The classical R/S analysis is quite robust against variations in the marginal distribution of the data. This is also true for data with infinite variance.

Calculating the Hurst Exponent H and the stability index α of the process innovations, the long-range dependence parameter d is obtained by

$$d = H - 1/2, \tag{5.23}$$

for finite variance $(\alpha = 2)$, and by

$$d = H - 1/\alpha, \tag{5.24}$$

for infinite variance $(\alpha < 2)$.

Long-range dependence occurs if d is greater than 0.

The R/S analysis is a nonparametric tool for examining long-memory effects. There is no requirement for the time series's underlying limiting distribution. In case of an underlying Gaussian process ($\alpha = 2$), a Hurst Exponent of H = 0.5 implies that there is no long-range dependence among the elements of the time series.

For $0.5 < H < 1$, a Gaussian time series is called *persistent*.[12] A persistent time series is characterized by long-memory effects. If long memory is present, the effects occur regardless of the scale of the time series. All daily changes are correlated with all future daily changes, and all weekly changes are correlated with all future weekly changes. The fact that there is no characteristic time scale is a primary property of fractal time series.

$0 < H < 0.5$ signals an *antipersistent* system for finite variance. Such a system reverses itself more frequently than a purely random one. At the first glance, it looks like a mean-reverting process. But this would actually require a stable mean, which is not the case in such systems.

[12]Peters (1994, p. 61 and chapt. 13).

5.3.4 The Modified Approach By Lo

Hurst's R/S statistic turned out to react sensitively towards short memory processes. Hence, Lo (1991) modified the classical R/S statistic, which shows robustness towards short-range dependence. Lo's statistic focuses solely on lag $n = N$, the length of the series.[13] Multiple lags are not analyzed, the statistic does not vary n over several lags $< N$.

Compared to the graphical R/S method, which delivers an estimate of the parameter H, Lo's modified statistic merely indicates the presence of long-range dependence, but does not deliver an estimate of the Hurst Exponent. The statistic performs a test of the hypotheses H_0: no long-range dependence.

Instead of the ordinary sample standard deviation S for normalization, there is an adjusted standard deviation S_q in the denominator. S_q effects the elimination of short-term memory to the statistic. As it is known that the R/S statistic responds very sensitively towards short-range dependence, the influence of short-range dependence may be offset by normalizing R with a weighted sum of short-lag autocovariances. To the variance S^2, Lo added weighted autocovariances up to order q. [14] His modified statistic $V_q(N)$ is defined by

$$V_q(N) = N^{-1/2} \frac{R(N)}{S_q(N)}, \tag{5.25}$$

$$\text{with } S_q(N) = \sqrt{S^2 + 2 \sum_{j=1}^{q} w_j(q) \hat{\gamma}_j}, \tag{5.26}$$

where $\hat{\gamma}_j$ is the autocovariance of order j for the observed time series. $w_j(q)$ is defined as

$$w_j(q) = 1 - \frac{j}{q+1} \text{ with } q < N. \tag{5.27}$$

The statistic $V_q(N)$ is applied for a hypothesis test. It checks whether the null hypothesis of the test can be rejected or not, given a certain confidence level. The two hypotheses are as follows:

- H_0: no long-range dependence present in the observed data, $0 < H \leq 0.5$.
- H_1: long-range dependence is present in the data, $0.5 < H < 1$.

[13] Teverovsky, Taqqu and Willinger (1998).
[14] Lo (1991)).

The statistic assumes a Gaussian process ($\alpha = 2$). In cases where the value of $V_q(N)$ lies inside the interval [.809, 1.862], H_0 is accepted since the statistic is in the 95% acceptance region. For $V_q(N)$ outside the interval [.809, 1.862], H_0 is rejected.

Lo's results are asymptotic assuming $N \to \infty$ and $q = q(N) \to \infty$.[15] However, in practice the sample size is finite and the value of the statistic depends on the chosen q. Thus, the question of what would be the proper value for q in order to perform the hypothesis test arises. Andrews (1991) has developed a data-driven method for selection of q:[16]

$$q_{opt} = \left[\left(\frac{3N}{2} \right)^{1/3} \left(\frac{2\hat{\rho}}{1 - \hat{\rho}^2} \right)^{2/3} \right], \tag{5.28}$$

where [ ] stands for the greatest integer smaller than the value in between. $\hat{\rho}$ is the first-order autocorrelation coefficient. Therefore, choosing Andrews's q assumes that the true underlying process is AR(1).

Critique of Lo's statistic. Lo's statistic is applied by calculating V_q for a number of lags q and plotting those values against q. The confidence interval for accepting H_0 at the 95% confidence level is plotted as well.

Simulations have shown that the acceptance of H_0 (and therefore the value of $V_q(N)$) varies significantly with q. Teverovsky, Taqqu and Willinger (1998) found that the larger the time series and the larger the value for q, the less likely H_0 is rejected.

Whereas Lo's statistic simply checks for the significance of long-range dependence, the graphical method of the classical R/S provides relatively good estimates of H.

For small q, the results of V_q usually vary strongly. Then a range of stability follows after the so-called extra short-range dependence has been eliminated, and the only effect measurable for the statistic would be long-range dependence.

Applying the statistic to Fractional Brownian Motion with $H > 0.5$, which is a purely long-range dependent process without short memory effects, V_q is expected to stabilize at very low values of q. Unfortunately this could not be confirmed in testing done by Teverovsky, Taqqu and Willinger (1998). Moreover, they demonstrate that, if q is large enough, the following holds for $V_q(N)$ and $q^{0.5-H}$:

$$V_q(N) \simeq q^{0.5-H}. \tag{5.29}$$

[15] Teverovsky, Taqqu and Willinger (1998).
[16] See Lo (1991).

For $H > 0.5$, V_q decreases with increasing q. Even for strongly fractional processes with time series containing 10,000 samples, Taqqu, Willinger, and Teverowsky found that, with increasing values for q, the probability that V_q will lie inside the H_0 95% confidence interval, thus accepting the null-hypothesis, grows. To mention three cases only: for q=500 and H = 0.9 the null-hypothesis (no long-range dependence) is accepted with 90% for Fractional Brownian Motion, with 92% for FARIMA(0.5, d, 0), and with 94% for FARIMA (0.9, d, 0).[17]

Lo's test is decidedly conservative in rejecting the null hypothesis. It works for short-range dependence, but in cases of long-range dependence it largely accepts the null hypothesis. The statistic proposed by Lo is certainly an improvement compared to the short-range sensitive classical R/S, but it should not be used alone, i.e. without comparing its results to other tests for LRD.

In practical applications, the issue of a proper choice of q remains. The value of Andrews's data-driven q_{opt} depends on the econometric model underlying the observed time series, but, the appropriate model is not known in advance. Andrews's choice bears the assumption that the time series obeys an AR(1) process.

A common way to assess long-range dependence used to be by looking at the rate at which the autocorrelations decay. With a Hurst Exponent H different from 0.5, the correlations are no longer summable. Such non-summability of autocorrelations used to be seen as a comfortable way of assuming long-range dependence. But there are pitfalls: if the underlying process is considered to follow a stable law with $\alpha < 2$, a second moment does not exist, which precludes the existence of autocorrelations.

It can be concluded, that when testing for long-range dependence, the application of a single technique is insufficient.

5.3.5 The Mansfield, Rachev, And Samorodnitsky's Statistic (MRS)

Long-range dependence means that a time series exhibits a certain kind of order over a long coherent period. Instead of pure chaos with no rule in the price movements of an asset, one can find periods of time with their sample mean significantly different from the theoretical mean. The stronger the long-memory effects in the time series, the longer an interval of the series whose mean deviates from the expected value.

Mansfield, Rachev and Samorodnitsky (1999) concentrate on this property of LRD-exhibiting time series. This property of LRD is valid regardless of the assumed underlying stochastic model.

[17] FARIMA(0.5, d, 0) means a fractional ARIMA process with an AR(1) coefficient of 0.5 and an MA(1) coefficient of 0.

The authors define a statistic that delivers the length of the longest interval within the time series, where the sample mean lies beyond a certain threshold. The threshold is set greater than the finite mean EX_i of the overall time series. Furthermore, the time series is assumed to follow a stationary ergodic process.

Expressed in mathematical terms, the statistic is defined as

$$R_n(A) = sup\{j - i : 0 \leq i < j \leq n, \frac{X_{i+1} + ... + X_j}{j - i} \in A\}, \tag{5.30}$$

which is defined for every $n = 1, 2, ...$ If the supremum is taken over the empty set, the statistic is defined to be equal to zero.

The set A is defined either as

$$A = (\theta, \infty) \text{ with } \theta > \mu, \tag{5.31}$$

or as

$$A = (-\infty, \theta) \text{ with } \theta < \mu, \tag{5.32}$$

where μ is the theoretical mean of the time series.

$R_n(-\infty, \theta)$ and $R_n(\theta, \infty)$ are interpreted as "greatest lengths of time intervals when the system runs under effective load that is different from the nominal load".[18] In the following, the examination is restricted to $R_n(\theta, \infty)$.

A theoretical method to examine a time series for long-range dependence would be the log-log plot of $R_n(\theta, \infty)$ versus n. In the case of long-range dependence, the slope of the plot would be expected to be greater than $1/\alpha$ with α as the tail index. However, α is not known in advance. For this reason, Mansfield, Rachev, and Samorodnitsky developed a statistic that does not rely on an a-priori tail index. They defined

$$W_n(\theta) = \frac{R_n(\theta, \infty)}{M_n}, \tag{5.33}$$

where $M_n = \max(X_1, ..., X_n)$ is the largest of the first n observations, $n \geq 1$. This statistic has a self-normalizing nature and, because of the denominator, it has the ability to compensate for the effects of the tail index α.

In case of short-range dependence, the ratio $W_n(\theta)$ approaches a weak limit as $n \to \infty$. In case of long-range dependence, R_n grows faster than M_n and the statistic diverges.

For visualization, the statistic $\theta W_n(\theta)$ is plotted against θ. Its limiting distribution is independent of θ. A difficult task is the selection of the proper range of θ, which has to be determined empirically by examining where the values for $\theta W_n(\theta)$ stabilize.

[18] See Mansfield, Rachev and Samorodnitsky (1999).

Index	Explanation
X0H0	High Yield 175
C8B0	Corporates C rated, cash pay
J0A3	AAA-AA rated Corporates, time to maturity 15 yrs
C0A0	US Corporate Master

Table 5.1: Explanation of the selected indices.

Index	No. of Observations	Starting Date	Ending Date
X0H0	3083	10—31—86	04—30—00
C8B0	3470	10—31—86	04—30—00
J0A3	2920	08—04—88	04—30—00
C0A0	3472	01—04—88	04—30—00

Table 5.2: Data sets used for testing LRD.

Once the value of the statistic is at least 19 for a certain θ, then long-range dependence is present at a significance level of 0.05.

5.4 Empirical Results: LRD In Credit Returns

For empirical examination of long-memory effects in daily credit return data, the returns of bond indices provided by Merill Lynch are chosen.[19] Four indices with time series of daily index returns have been selected (between January 1988 and April 2000). The number of available observations for each time series ranges from 2920 to 3472. Each index represents a number of bonds with similar properties (see explanation in table 5.1). As the analysis of long-memory effects requires large data samples, an important criterion for the selection of an index was the available sample size. The sample sizes are listed in table 5.2.[20]

Three different methods for estimating the self-similarity parameter H are applied. Two methods perform a hypothesis test regarding the presence of LRD. As explained before,

- (i) the Aggregated Variance method,
- (ii) the Absolute Values of Aggregated Series method,
- (iii) the classical R/S analysis developed by Mandelbrot and Wallis,

[19] The time series were obtained via Bloomberg's Index Section.
[20] The results are shown in Martin, Rachev and Schwartz (2002).

- (iv) Lo's modified R/S statistic,
- (v) the statistic of Mansfield, Rachev, and Samorodnitsky (MRS).

All these methods have been implemented with Matlab 5.3. Methods (i) - (iii) provide an estimate of the Hurst Exponent H. Method (iv) examines whether the null hypothesis *no long-range dependence* has to be accepted or rejected at a given confidence level. Method (v) is also a hypothesis test, however, in contrast to Lo's test, it works independently of the tail index.

When testing the index returns for long-range dependence, the daily changes of the index log prices are calculated as follows:

$$r_t = log(p_t) - log(p_{t-1}). \tag{5.34}$$

The results of the methods Aggregated Variance and Absolute Values of the Aggregated Series. For methods (i) and (ii), the values of each statistic are plotted over m (number of elements in each block), with m ranging from 10 to 40. Finally, the slope of the data points is determined in order to obtain H. The values for H are set out in table 5.3. Both methods (i) and (ii) find Hurst Exponents greater than 0.5 for all observed indices. Thus, under the Gaussian assumption, the underlying processes are long-memory processes. X0H0 and J0A3 show strong LRD, whereas C8B0 and C0A0 have a weaker long memory.

Index	H for Aggreg. Variance	H for Abs. Values of Aggreg. Ser.
X0H0	0.7632	0.7596
C8B0	0.5527	0.5511
J0A3	0.8070	0.8022
C0A0	0.5856	0.5838

Table 5.3: The results for Aggregated Variance and Absolute Values of the Aggregated Series.

The results of classical R/S and Lo's statistic. As there are only about 3000 observations for each time series, the data set is not divided into several blocks for the classical R/S statistic. Thus, it makes sense to choose K = 1.

The results of classical R/S and the values of Lo‘s statistic V_q (for q the range of 1...50 is chosen) are presented in table 5.4. Both the log(R/S)-log(n) and the $V_q - q$ graphs are plotted for the observed indices X0H0, C8B0, J0A3, and C0A0 (see figures 5.1, 5.2, 5.3, 5.4 for classical R/S , and figure 5.5 for Lo's test). The second column in table 5.4 presents the Hurst Exponent estimated with the R/S statistic. In the third column the table depicts the intervals in which the values

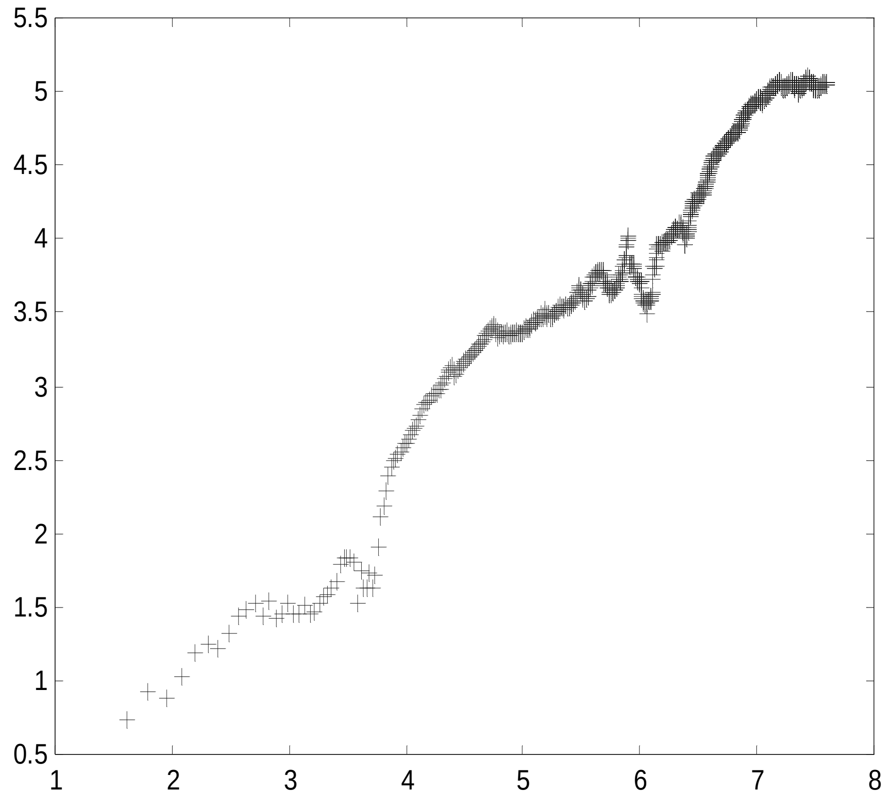

Figure 5.1: Plot of $log(RS) - log(n)$ for X0H0.

of Lo's V_q are located for $q = 1...50$. The fourth column provides the optimal lag q, determined by Andrews's data-driven method.[21] The results of the R/S statistic are similar to those obtained by the Aggregated Variance method and the Absolute Values of the Aggregated Series. The time series of X0H0 and J0A3 exhibit strong LRD indicated by their Hurst Exponent H. This is supported by the result of Lo's test that rejects the null-hypothesis *no LRD* at the 95% level. However, for C8B0 and C0A0, the Hurst Exponent already appears in the area of antipersistence. Another interesting finding is that for C0A0 - which has the lowest value for H - the sample autocorrelation of order 1 is negative. Therefore, the optimal q for C0A0 cannot be calculated.

[21] See Lo (1991).

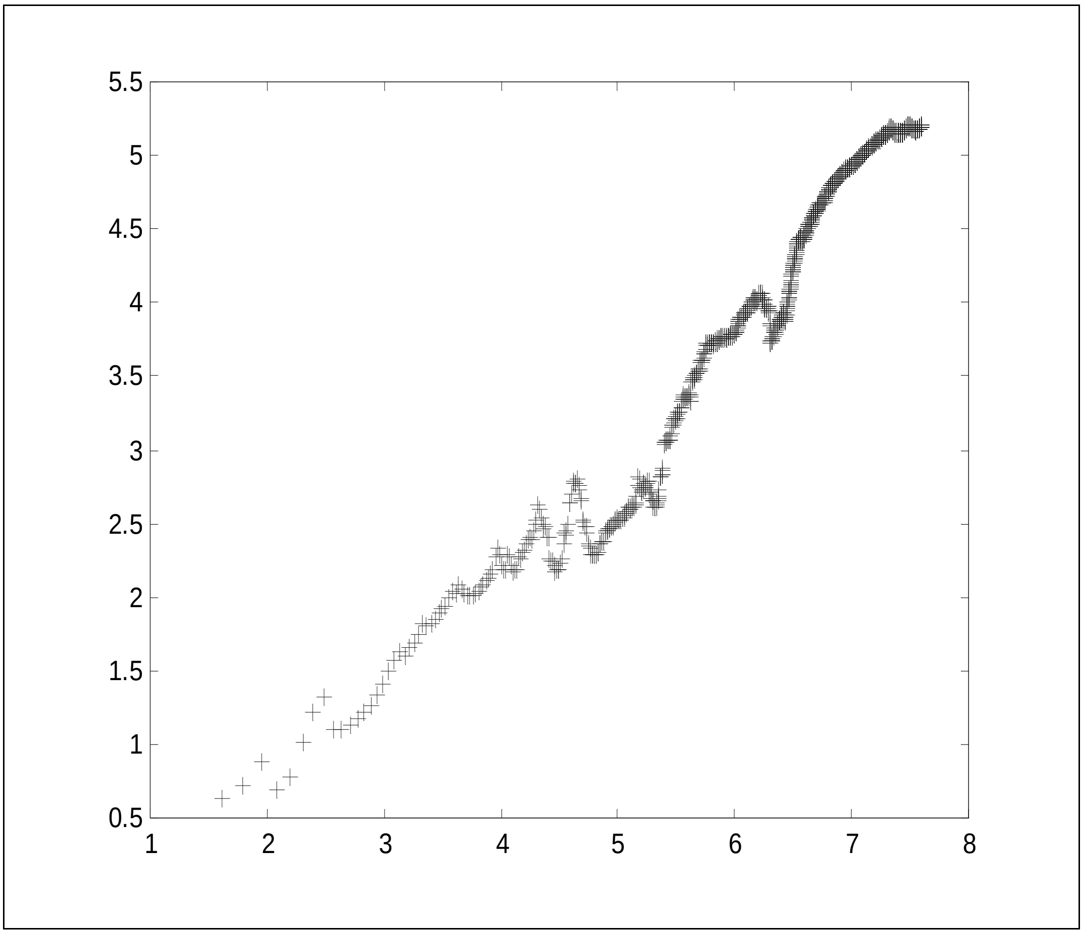

Figure 5.2: Plot of $log(RS) - log(n)$ for J0A3.

The results of the Mansfield, Rachev, and Samorodnitsky's (MRS) statistic. Figures (5.6) - (5.9) show the plots of $\theta W_n(\theta)$ over the range of θ. For the time series of the index X0H0, the statistic $\theta W_n(\theta)$ increases linearly with θ in the range of $[0.5e^{-4}, 3.5e^{-4}]$ (the empirical mean of the overall series is $0.497e^{-4}$). The value of $\theta W_n(\theta)$ reaches levels of about 19 and then declines until it stabilizes at a level of about 1 (see figure 5.6). This result clearly indicates the presence of LRD. The presence of long memory is significant at the 0.05 level once the value of the statistic is at least 19. Thus, the MRS statistic supports the LRD hypothesis for X0H0. Lo's statistic and classical R/S also indicate long-range dependence for the index X0H0, but this was based on the assumption that the underlying process of the time series follows a Gaussian law, i.e. that $\alpha = 2$. However, the MRS statistic is independent of α.

The second bond index that exhibits strong LRD in its returns with the former tests, was the J0A3-index (C rated corporates). Its empirical mean is $-3.2e^{-4}$.

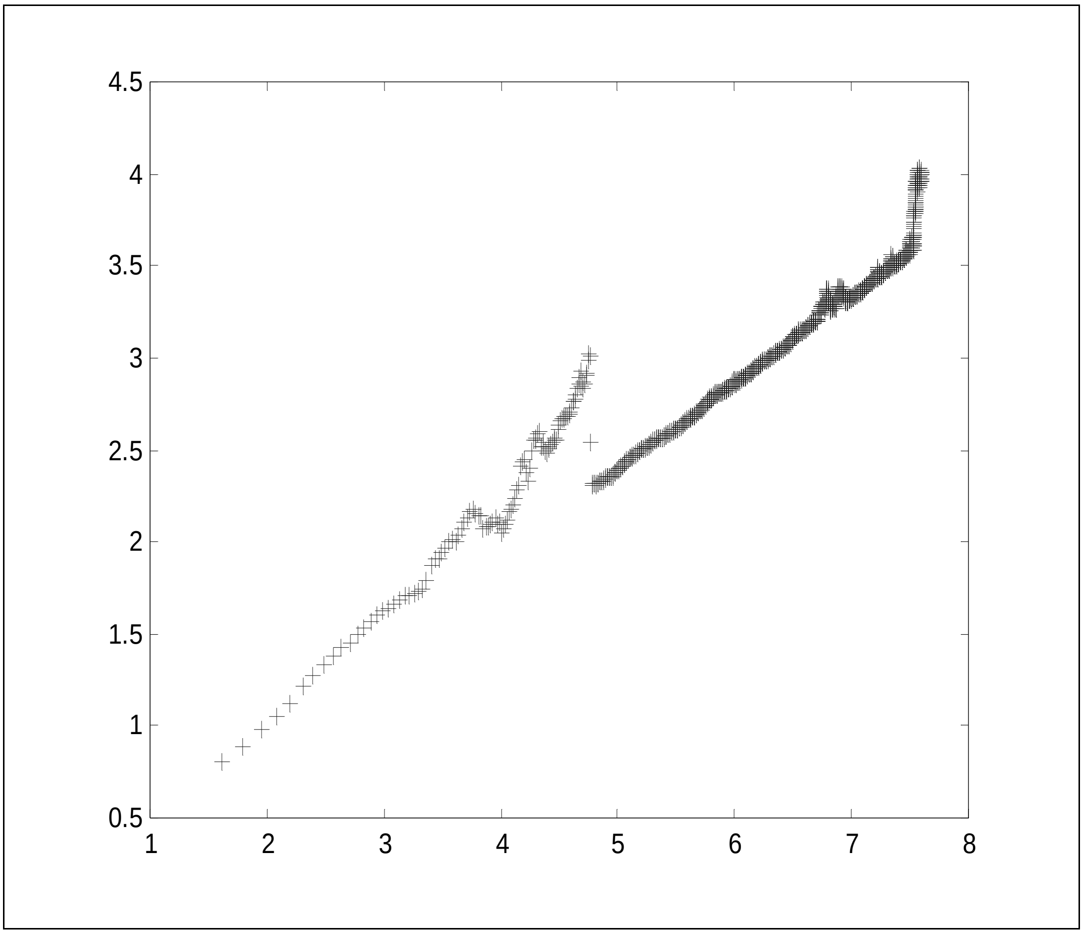

Figure 5.3: Plot of $log(RS) - log(n)$ for C0A0.

$\theta W(\theta)$ increases sharply for $\theta \in [0, 6.5e^{-4}]$ to a value of about 15, finally dropping to a level of about 1. Thus, the hypothesis of long-range dependence can also be confirmed for the J0A3 series as the MRS statistic also exhibits significant values (see figure 5.7). However, the significance is not as great as for the X0H0 series.

The returns of the two other indices, C0A0 and C8B0, do not exhibit long-range dependence with the $\theta W(\theta)$ statistic, and this is consistent with the results of the previously applied tests. The returns of the C8B0 index show a higher probability for the LRD-hypothesis than the returns of C0A0; yet, both are not significant. Thus, for both indices C0A0 and C8B0, there is no significant indication for long-range dependence with the MRS statistic (see figures 5.8 and 5.9).

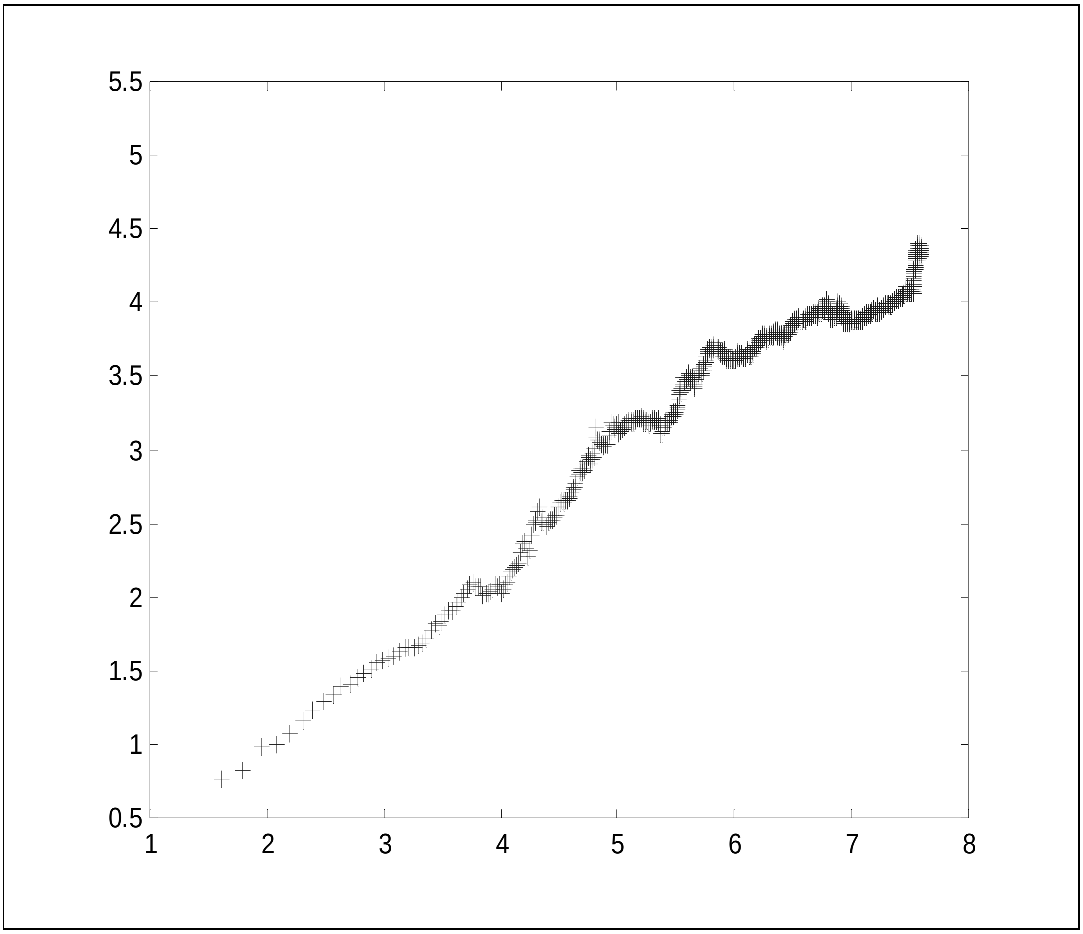

Figure 5.4: Plot of $log(RS) - log(n)$ for C8B0.

5.5 Conclusion

The idea of the Fractional Market Hypothesis is based on the notion that price movements do not obey a purely random walk, instead showing certain patterns of long-memory behavior.

A sign of long memory or long-range dependence (LRD) is the burstiness of plotted time series. Long-range dependence is characterized by hyperbolically decaying autocorrelations and the property that large (small) representations are more likely to be followed by large (small) representations than small (large) representations.

While three of the five tests that have been applied to the corporate bond indices measure the Hurst Exponent, the other two are hypothesis tests which check significance of LRD.

Applying the Aggregated Variance and Absolute Values of Aggregated Series

	Classical RS	Lo's Statistic	
Index	Fitted H	Range of $V_q(q = 1...50)$	Optimal q (Andrews)
XOHO	0.7579	[1.74, 3.44]	11
C8B0	0.4874	[1.33, 1.40]	6
J0A3	0.9213	[2.04, 4.29]	10
C0A0	0.4493	[1.23, 1.40]	-

Table 5.4: Results for the classical R/S statistic and Lo's test.

methods, all four time series analyzed exhibit a Hurst Exponent H greater than 0.5, which indicates long-range dependence under the Gaussian assumption. For two of the four credit return series, the modified R/S statistic developed by Lo confirms LRD as significant. This is remarkable because Lo's test tends to confirm the null hypothesis *no LRD* for large sample sizes and increasing lag q, even when the actual process is strongly long-range dependent.[22] Also allowing infinite variance ($\alpha < 2$), the MRS statistic is applied. It analyzes a process for LRD without relying on the tail index. For the X0H0 and J0A3 series, which have been confirmed for LRD by Lo's test, the MRS statistic $\theta W(\theta)$ also indicates significant long memory.

The most distinct result of the LRD studies in this chapter is that long-range dependence in credit returns is also found to be *significant in combination with the non-Gaussian stable assumption*.

The examinations here have only focused on the returns. However, for other financial series such as stock prices, LRD has also been discovered in the trading time process as demonstrated by Marinelli, Rachev, Roll and Göppl (1999).

The use of bond indices for the empirical examination instead of individual bonds is advantageous in two respects:

- First, each index incorporates numerous bonds of a certain market segment. Thus, the results obtained can then be considered a widespread phenomenon. If only a small number of bonds within the observed indices were to exhibit such an effect, it would probably fade away.
- Second, LRD analysis requires large samples, which are more readily available for indices than for single bonds.

Chordia, Roll, and Subrahmanyan's findings (2000a) on the liquidity anomaly are another interesting aspect. They maintain that illiquid periods - when discovered by the agents - will further reduce liquidity. Similarly, rising markets attract more investors. Such a phenomenon is also characteristic of both GARCH and long-memory processes.

[22] See Teverovsky, Taqqu and Willinger (1998).

Finally the conclusion can be drawn that the issue of long memory may not be neglected for time series of credit returns. The increments of the underlying stochastic process are not i.i.d.

The proven LRD in the time series of credit returns and the demonstration that the distribution of credit returns is better captured with stable non-Gaussian models offers a powerful tool for generating accurate forecasts of VaR, especially for longer horizons.

These tools will play a vital role in the following chapters.

The next chapter discovers another phenomenon observed for credit data: log prices of corporate bonds with different credit qualities exhibit common long-term behavior for given time to maturity. Such behavior is called cointegration.

Chapter 6 examines the phenomenon. Chapter 7 introduces techniques suitable for capturing this long-term behavior. Chapter 8 builds the model.

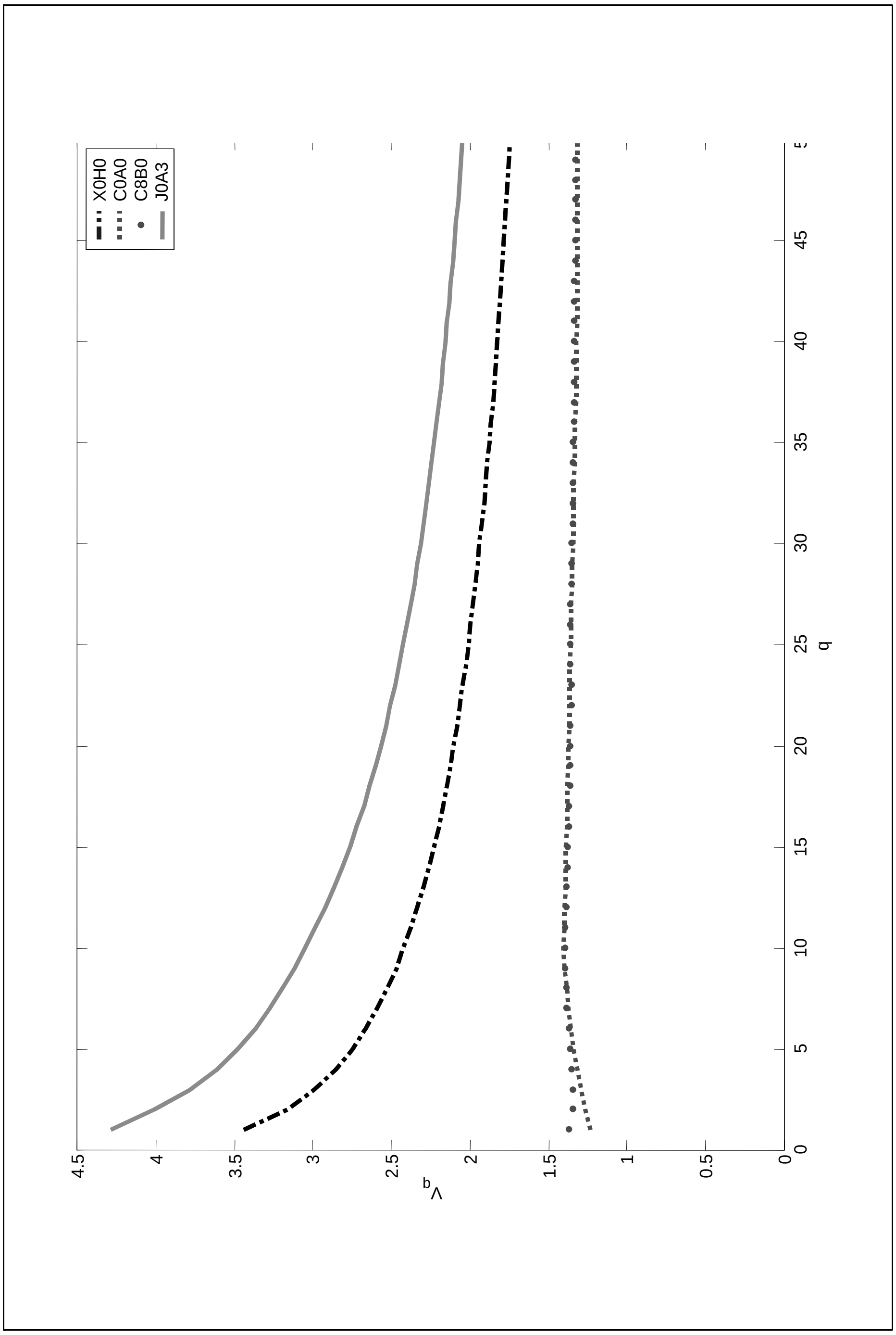

Figure 5.5: The plots of Lo's statistic for X0H0, J0A3, C0A0, and C8B0

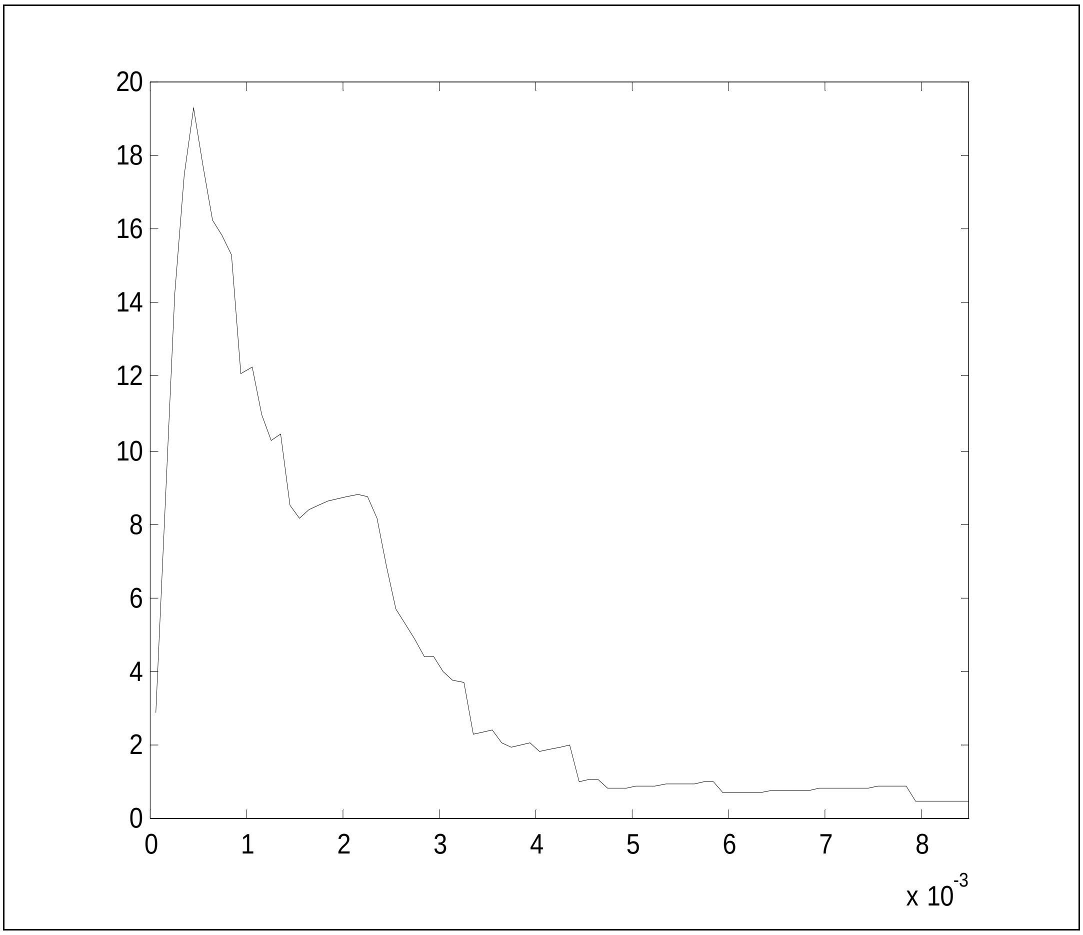

Figure 5.6: Plot of $\theta W(\theta) - \theta$ for X0H0.

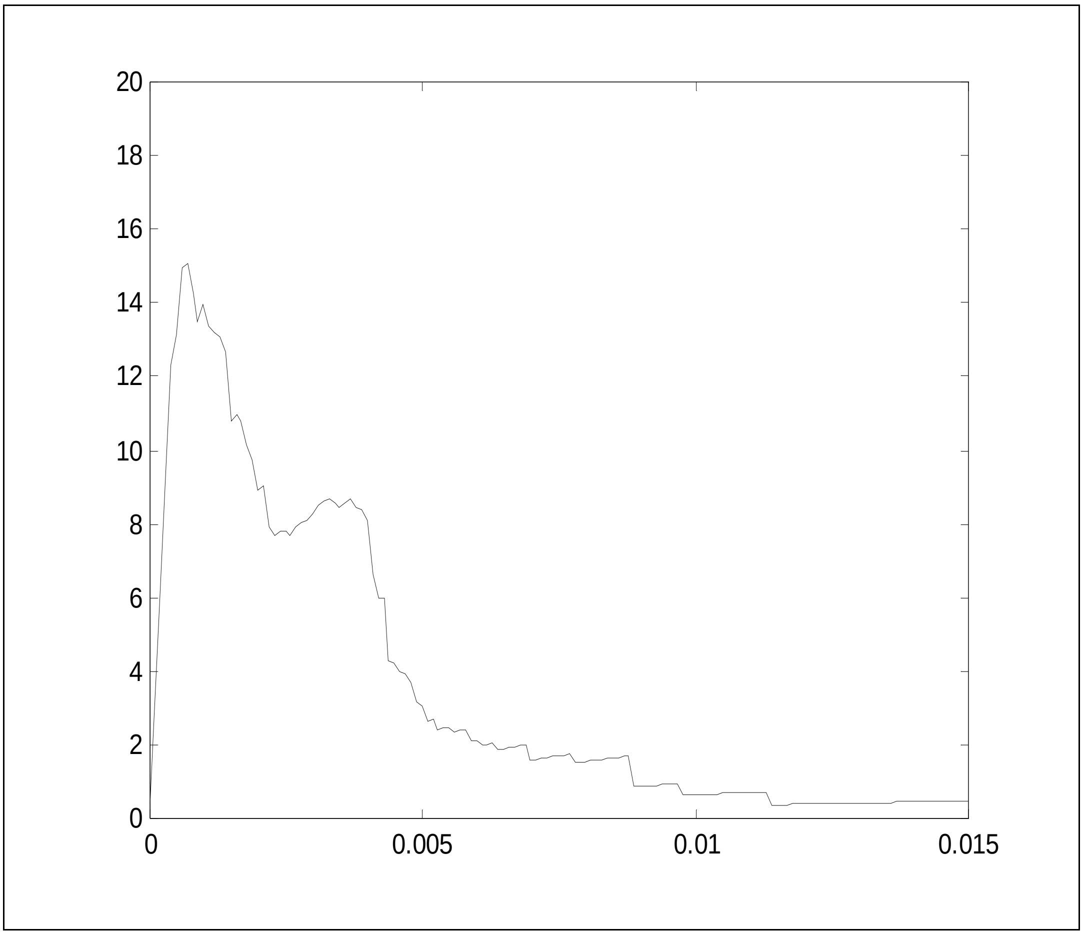

Figure 5.7: Plot of $\theta W(\theta) - \theta$ for J0A3.

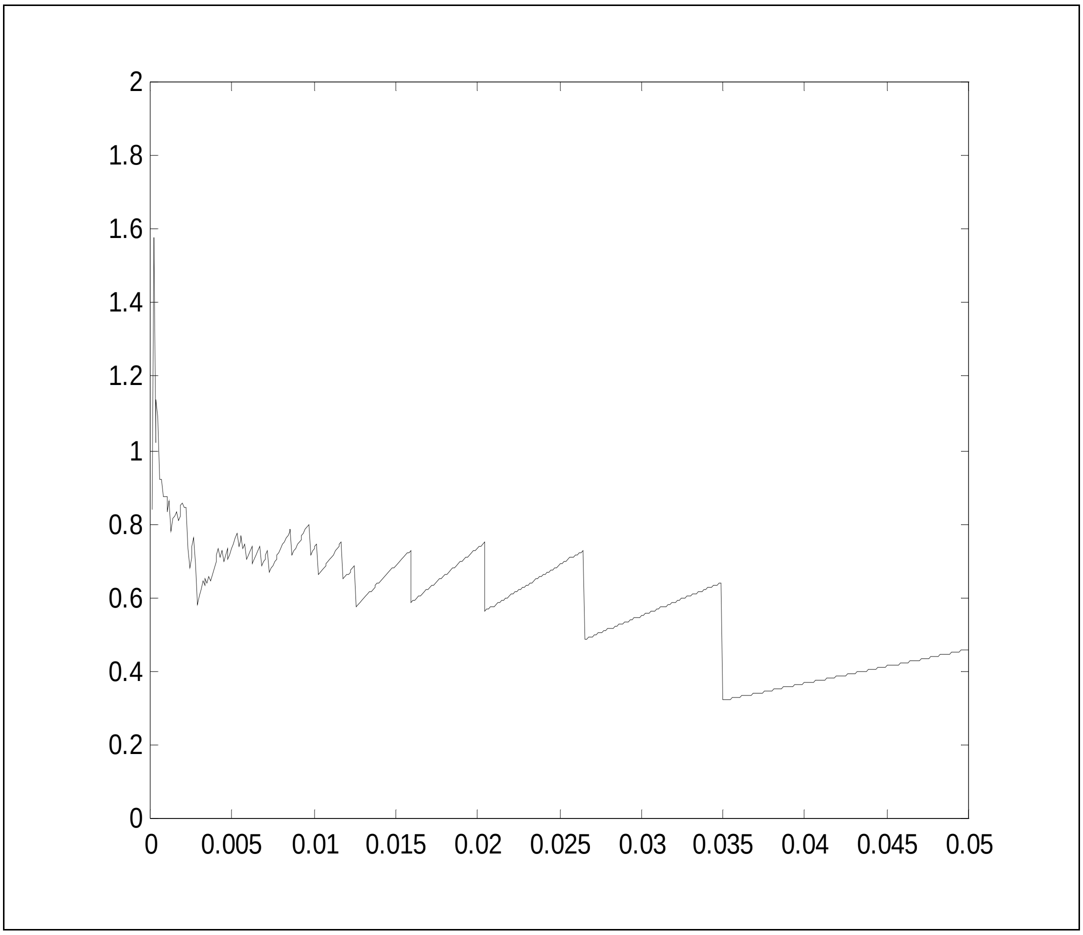

Figure 5.8: Plot of $\theta W(\theta) - \theta$ for C0A0.

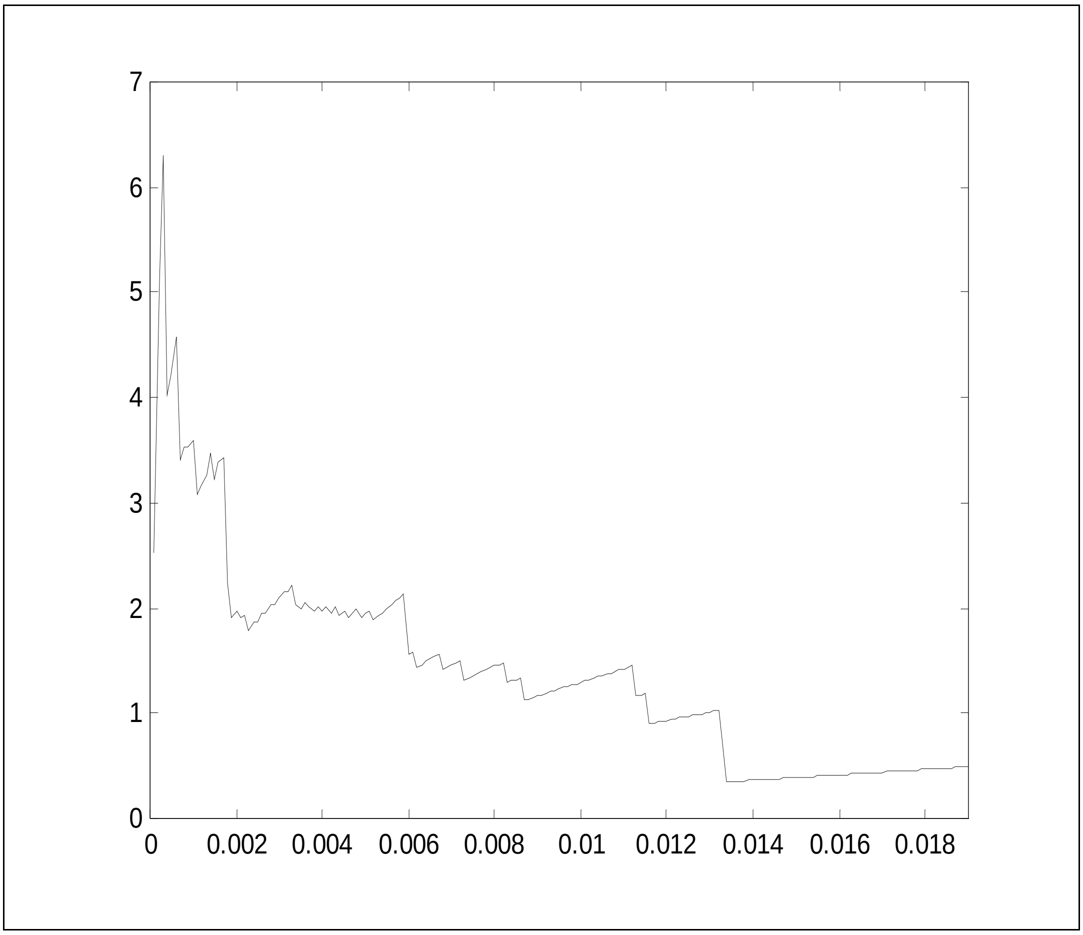

Figure 5.9: Plot of $\theta W(\theta) - \theta$ for C8B0.

Chapter 6

Modeling The Returns Of Different Credit Ratings

6.1 Empirical Evidence: New Phenomena In Credit Data

In section 3.2, a one-factor model for bonds bearing credit risk has been introduced. The returns of the individual bonds are described by the following components:

- the returns of a corresponding treasury bond.
- the movement in credit spread.

The model is based on the stable assumption which shows a much better fit and forecast results for VaR.

So far, dependence between the returns of different corporate bonds has been captured by modeling (i) the dependence between the corresponding treasury returns and (ii) the dependence between the disturbance terms (credit spread changes). This accords the model in chapter 3 the ability to capture portfolio effects.

However the direct modeling of individual corporate bonds' returns is not optimal for a number of reasons:

1. Aside from short-term dependence, there are also long-term relationships between the log price paths of different credit qualities, given equal maturity.
2. The model introduced in section 3.2 works directly with the observed price data of the corporate bonds. However, so far this model has not been able to take into account that changing time to maturity might have an impact

on the model's parameters. This issue can be avoided by working with a set of defined, constant times to maturity.

In this chapter, for the first time, a *stable cointegrated vector-autoregressive* approach is proposed as a common framework to describe corporate bond returns of different credit qualities with fixed time to maturity. A set of fixed times to maturity has been defined. For each time to maturity, a cointegrated vector-autoregressive model is built. This model incorporates the requirements resulting from the essence of the above items 1. and 2. This approach enables both short- and long-term properties to be captured.

The chapter begins with preliminary analyses of the data as a prerequisite for the model building process. Then the theory of cointegrated processes is explained. The following chapter 7 goes into the theory of cointegrated vector-autoregressive (VAR) models and presents the methods used to obtain their proper specification. In chapter 8, the suggested model is specified and fitted with empirical data. Chapter 9 discusses the behavior of the volatility of the credit returns over time and its implications on VaR. As a result, the performance of several multivariate volatility models based on the stable assumption is compared for VaR.

Prices of corporate bonds with different credit ratings show similar behavior over time. In particular those corporate bonds with equal maturity exhibit certain relationships in their price paths. It is the objective of the following part of the thesis to analyze and describe such behavior for zero-coupon corporate bonds over all credit ratings and maturities.

For risk management purposes, it is essential to determine the potential daily price changes of a bond with a given credit rating and a given time to maturity. Instead of modeling the daily yield curves, it is advantageous to focus directly on the daily returns as risk managers are interested in the calculation of a credit portfolio's VaR. Figure 6.1 presents a four-year history of the average log prices for US dollar-denominated bonds maturing in 10 years. It is displayed for all credit ratings from AAA to B.

The log prices of different corporate credit ratings are largely determined by common movements and similar trends in the long term. Furthermore one finds that the returns of the corporate credit ratings are to a great degree driven by treasury bond returns of equal maturity. The higher the credit rating, the greater is the influence of the treasury bond returns. Regressing the corporate bond returns over the treasury returns, the regression explains roughly 40% of the corporate bond returns for credit quality B and about 72% of the returns for credit quality AAA.[1] Figure 6.2 depicts the historical 10-year treasury returns and the 10-year returns of rating grades AAA - B over an observation period of four years. As the

[1] Data is available for corporate bonds with credit ratings from AAA to B.

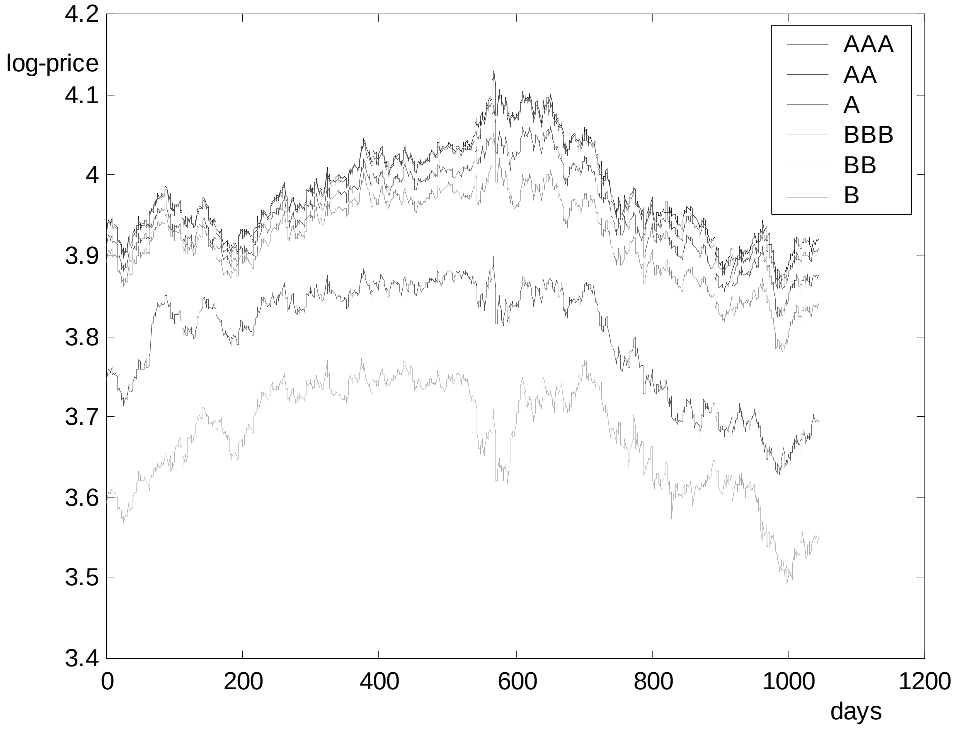

Figure 6.1: Daily average log prices for corporate rating grades AAA - B with 10-year maturity from 8/96 to 7/00, US market.

residuals of the regressions indicate strong dependence, the application of seemingly unrelated regressions may be a feasible for a vector-autoregressive system of six equations.

Rating	R^2
AAA	72.3 %
AA	73.3%
A	73.4%
BBB	72.3%
BB	52.2%
B	41.0%

Table 6.1: Fraction of corporate bond returns explained by the regression over treasury returns (10-year maturity).

The obtained regression parameters are listed in table 6.1. Further examination of the residuals shows that these are heavy-tailed. The plot of the residuals also illustrates signs of clustering volatility. This means volatility varies over time with periods of large and periods of small volatility. This phenomenon is examined later in this work. Examination of the log prices in figure 6.1 above shows that they follow a common long-term trend. Thus, the idea is now to capture such a behavior with a system of equations through the application of a cointegrated vector-autoregressive (VAR) system. The dependent variables on the left-hand side of the VAR are the returns of the corporate zero-bonds, each belonging to a different credit rating. The riskless credit returns enter the model as an exogeneous variable. The cointegrated VAR will finally be represented by an *error correction model (ECM)* whose mechanism is introduced in the following sections.

Both the impressions gained from figures 6.1, and 6.2 and table 6.1 indicate an improvement of information by applying such a systematic multivariate approach compared to the approach presented in section 3.2 (separate modeling of individual bonds).

But before the cointegrated VAR model for the credit returns is specified in depth, the concept of cointegration is introduced with a special focus on the case when the innovations of the cointegrated model follow a stable law. Moreover, the functionality of error correction models is explained. Testing procedures for detection of cointegration are discussed. A section on unit root theory provides the theoretical background. Section 6.2.5 discusses unit-root theory and cointegration under and the stable non-Gaussian assumption.

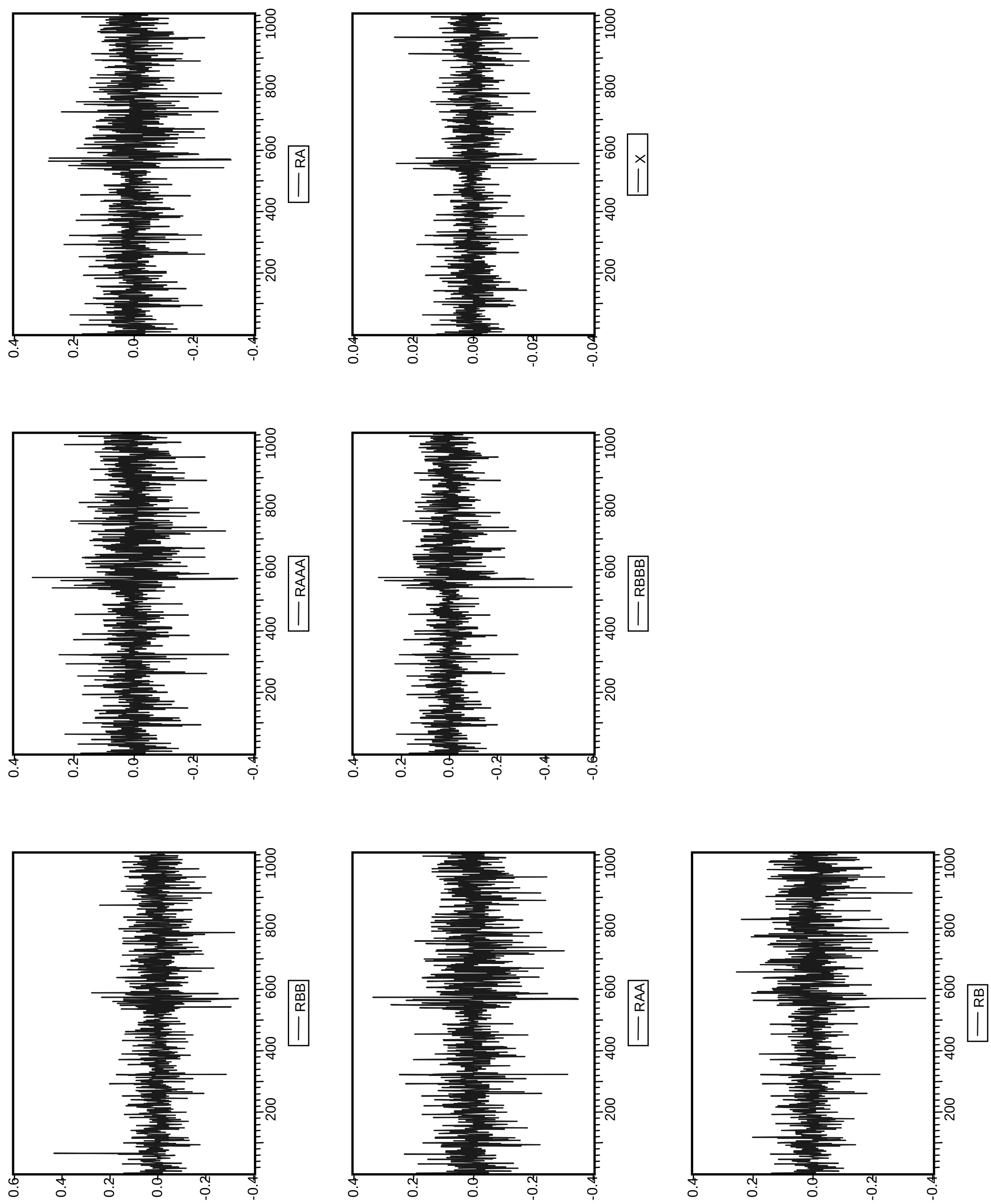

Figure 6.2: Daily average returns for corporate rating grades AAA - B with 10-year maturity from 8/96 to 7/00, US market. X denotes the treasury bond returns.

6.2 The Concept Of Cointegration

This section reviews the general concept of cointegrated series for the bivariate case, explains the functioning of error correction models, and presents tests for cointegration.

Generally, trended data is a major problem of empirical econometrics. Trends can cause spurious regressions. Goodness-of-fit measures might lead to doubtful results. In financial econometrics, most variables are subject to some kind of trend, stochastic or deterministic. As a result, *differenced* or *log-differenced* variables, e.g. financial returns, are preferred for prediction models in order to deal with stationary variables. However, rendering variables stationary means that long-term properties cannot be described. The concept of cointegration was developed in order to obtain models demonstrating both long- and short-term properties and stationarity in all variables of the model. The assumption behind this is that the observed time series follow a long-term equilibrium. Cointegration means that there exists a linear combination of non-stationary (trended) variables that is stationary itself. Series exhibiting such properties are called *cointegrated series.*[2] The following subsection presents an example of cointegration for the case of two variables.

6.2.1 A Case With Two Variables

Assuming there is a bivariate process generating y_t and x_t with

$$x_t = c + x_{t-1} + \epsilon_t \ , \ \epsilon_t \text{ i.i.d. with } N(0, \sigma_1) \ , \tag{6.1}$$

and a linear combination of x_t and y_t such that

$$y_t = a + x_t + u_t. \tag{6.2}$$

Suppose, u_t is generated by

$$u_t = f + \rho u_{t-1} + \nu_t \text{ with } \nu_t \sim N(0, \sigma^2) \ . \tag{6.3}$$

If u_t results in non-stationarity, then both series x_t and y_t are drifting apart. In this case, the series y_t and x_t are *not* cointegrated. If u_t is stationary then the series are cointegrated. The formal definition of cointegration was developed by Engle and Granger (1987) for the two-variable case:

Two time series, x_t and y_t are said to be cointegrated of order (d, b), written as x_t, $y_t \sim CI(d, b)$, with $d \geq b \geq 0$, if

[2] Charemza and Deadman (1997).

1. Both series are integrated of order d, $x_t \sim I(d)$ and $y_t \sim I(d)$. That is these time series have to be differenced d times in order to become stationary.

2. There exists a linear combination of these variables which is integrated of order $d-b$. $[\beta_1, \beta_2]$ is the vector of parameters of the linear combination. It is called the *cointegrating vector*.

To achieve stationarity, a time series may have to be differenced more than once.

The order of integration is defined as follows:[3] A non-stationary series which can be transformed into a stationary series by differencing d times is said to be integrated of order d. It is conventionally denoted as $X_t \sim I(d)$. For example, if $X_t \sim I(2)$, the time series has to be differenced twice in order to obtain stationarity:

$$\Delta^2 X_t = \Delta(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2} \tag{6.4}$$

Standard tests to detect the order of integration are the Dickey-Fuller test or the augmented Dickey-Fuller (ADF) test.[4] Dickey-Fuller tests are a so-called *unit root tests* which check the null hypothesis that the observed time series has a *unit-root*. If the null hypothesis is confirmed, then the process is non-stationary and exhibits a trend. An explanation of the term *unit root* and a description of the Dickey-Fuller testing procedure are given in section 6.2.4.

6.2.2 Error Correction Models

A vector error correction model (VECM) is a restricted VAR incorporating cointegration restrictions in its specification. The restrictions are derived from the cointegrating equations which describe the long-run equilibrium relationship among the involved variables. The simplest case of cointegration is a two variable system with the cointegrating equation $y_{2,t} = \beta y_{1,t}$ and no lagged difference terms. The error correction for this case is

$$\Delta y_{1,t} = \gamma_1(y_{2,t-1} - \beta y_{1,t-1}) + \epsilon_{1,t}$$

$$\Delta y_{2,t} = \gamma_2(y_{2,t-1} - \beta y_{1,t-1}) + \epsilon_{2,t}. \tag{6.5}$$

If the cointegrating equation contains a constant, then there is a trend in the error correction model (ECM). The ECM links the realized value $y_{2,t}$ to its target

[3]See Engle and Granger (1987).
[4]Dickey and Fuller (1979); Dickey and Fuller (1981); Said and Dickey (1984) .

value (determined by the cointegrating relation) $y^*_{2,t} = \beta' y_{1,t}$. In case of a vector-autoregressive system, each equation of the ECM will also include terms of lagged Δy. Moreover, ECMs can include trends.

6.2.3 Testing For Cointegration

For conventional cointegration, the elements of y_t are assumed to be integrated of order 1 (denoted by I(1)) and there exists a linear combination among them that is integrated of order 0 (denoted by I(0)).[5]

$$\beta' y_t = u_t \text{ , with } u_t \sim I(0). \tag{6.6}$$

β is the *cointegrating vector*. $\beta' y_t = u_t$ is the *cointegrating regression*. If there exist two vectors β_1 and β_2 such that $\beta_1' y_t = u_{1t}$ and $\beta_2' y_t = u_{2t}$ are both I(0), then any linear combination of the two vectors is again a cointegrating vector. This is due to the fact that linear combinations of I(0) variables are again I(0). Thus, in case of more than two variables, the cointegration vector is no longer unique. The earliest cointegration test stems from Granger and Engle. It estimates the cointegration relation by ordinary least squares (OLS) estimation and applies a unit-root test to u_t in order to check for stationarity. Critical values for this unit-root test are tabulated.[6] Since the test focuses on u_t, it is called a *residual-based* test.

6.2.4 Unit-Roots And Integrated Processes

Before analyzing the cointegration relations between two or more variables, the individual orders of integration of the endogeneous variables have to be checked.

A time series y_t is said to have an autoregressive unit-root if it can be described by

$$y_t = d_t + z_t \text{ ,} \tag{6.7}$$

with

$$z_t = \rho z_{t-1} + \epsilon_t \text{ and } \rho = 1 \text{ ,} \tag{6.8}$$

where ϵ_t is stationary and d_t is a deterministic component. With $\rho = 1$, Δz_t is stationary, and Δy_t is stationary around the change of the deterministic component

[5] $CI(d, b)$ with $d = 1$ and $b = 1$. $d - b = 0$.

[6] See Engle and Granger (1987).

d_t. In this case, y_t is said to be integrated of order 1, i.e. I(1), and the stationary series, Δz_t and Δy_t, are said to be $I(0)$.[7]

With a unit root, i.e. an autoregressive parameter $\rho = 1$, the variance of process y_t increases over time. Thus, with an autoregressive parameter $\rho = 1$, the OLS estimator $\hat{\rho}$ no longer has an asymptotic normal distribution.[8] This also means that t-type or F-type tests based on an OLS estimator $\rho = 1$ do not follow the conventional t- or F-distribution.

By differencing a trended original variable y_t, obtaining $\Delta y_t = y_t - y_{t-1}$, a trend can often be removed. Sometimes the original variable has to be differenced more than once.

Most publications that analyze the asymptotic behavior of OLS estimates for $\hat{\rho}$ deal with the AR(1) model, $y_t = \rho y_{t-1} + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$.

Taking the autoregressive process of order 1, the transformed model is

$$\Delta y_t = (\rho - 1) y_{t-1} + \epsilon_t. \tag{6.9}$$

The hypothesis $H_0 : \rho = 1$ is tested against $H_1 : |p| < 1$ under the application of the t-statistic for $\hat{\rho}$:

$$t_{\hat{p}} = \frac{\hat{\rho} - 1}{\hat{s}_p}, \text{ where} \tag{6.10}$$

$\hat{\rho} - 1$ is the OLS estimate of the transformed model. $\hat{s}_p$ is the usual estimate for the standard deviation of the OLS estimate.

$$\hat{s}_p^2 = s^2 \left(\sum_{t-1}^{T} y_{t-1}^2 \right)^{-1}, \tag{6.11}$$

with s^2 as the estimate for the variance of the error term ϵ_t.[9] It has to be noted that $t_{\hat{p}}$ does not have the usual t- or asymptotic standard normal distribution. The relevant values are tabulated in Dickey and Fuller (1979).

Modifying equation (6.9) by adding a constant term yields

$$\Delta y_t = v + (\rho - 1) y_{t-1} + \epsilon_t. \tag{6.12}$$

Even if v is assumed to be zero, the t-statistic under the null hypothesis $\rho = 1$ has a different limiting distribution than in the former case since v has to be

---

[7] Annotation: Assuming y_t would be stationary, I(0), then Δy_t would be a moving-average unit root. MA unit roots arise from differencing stationary time series which is known as overdifferencing.

[8] Maddala and Kim (1998).

[9] Lütkepohl (1994).

estimated in addition to ρ. The relevant values are also in Dickey and Fuller (1979).

Further generalization is achieved by adding a deterministic trend to the model:

$$\Delta y_t = v + \xi t + (\rho - 1)y_{t-1} + \epsilon_t. \quad (6.13)$$

Here the relevant t-statistic of the null hypothesis $\rho = 1$ again has a different limiting distribution, which is, for example, tabulated in Dickey and Fuller (1979) or Hamilton (1994).

However, the processes in (6.9), (6.12), and (6.13) might not always be relevant for practical purposes. Sometimes, the actual process is more complex and has an autoregressive order greater than 1. Therefore, the Dickey-Fuller tests were made applicable also for AR processes of order $p > 1$. Such an autoregressive process

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \epsilon_t \ , \quad (6.14)$$

results in the transformation

$$\Delta y_t = -(1 - \alpha_1 - ... - \alpha_p)y_{t-1} + \pi_1 \Delta y_{t-1} + ... + \pi_{p-1}\Delta y_{t-p+1} + \epsilon_t. \quad (6.15)$$

Then, for the null hypothesis $H_0 : y_t$ is $I(1)$, it has to be confirmed that $1 - \alpha_1 .... - \alpha_p = 0$. H_0 can be checked with the t-statistic for the coefficient of y_{t-1} in combination with the Dickey-Fuller distribution.[10] This test is called augmented Dickey-Fuller (ADF).[11] This is the test that is commonly applied to examine stationarity for a wide range of AR-type processes. It may even be applied for ARMA processes, etc.

However, when testing for unit roots, merely applying available tests is not appropriate. Instead, a thorough test strategy considering pre-information should be developed.[12]

6.2.5 Unit Roots In The Stable Case

This section addresses unit root analysis under the assumption that the variables follow a stable law, based on the recognized equations introduced in the previous section. Following Rachev and Mittnik (2000, chapt. 15), the innovations of the processes obey a stable distribution and their time series are a Lévy process with

[10] Lütkepohl (1994).
[11] Said and Dickey (1984); Dickey and Fuller (1981); Dickey and Fuller (1979).
[12] Elder and Kennedy (2001).

strictly stable increments. The three relevant regression equations that are selected for the standard Dickey-Fuller test are listed below:

$$\Delta y_t = \beta y_{t-1} + u_t \text{ (Case I)} \tag{6.16}$$
$$\Delta y_t = \mu + \beta y_{t-1} + u_t \text{ (Case II)} \tag{6.17}$$
$$\Delta y_t = \mu + \tau t + \beta y_{t-1} + u_t \text{ (Case III)} . \tag{6.18}$$

For the first two regressions, the data-generating process (DGP) is of the form $y_t = y_{t-1} + u_t$. For the third equation, the DGP is $y_t = \mu + y_{t-1} + u_t$. The following analysis covers the cases I and II only.

The asymptotic behavior of the estimate for β and the t-statistic is examined for the first two of the above equations. The stable distribution which is assumed for the error terms is restricted to $1 < \alpha < 2$, and the u_t are in the domain of a strictly stable law. Now the asymptotic behavior of $\hat{\beta}$ and the t-statistics is analyzed under the null hypothesis $\beta = 1$. For the first-order autoregressive process $\Delta y_t = \beta y_{t-1} + u_t$ (case I), with $y_0 = 0$ and the u_t i.i.d., OLSE yields

$$\hat{\beta} = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2}. \tag{6.19}$$

Moreover, the t-statistic $t_{\hat{\beta}} = \frac{\hat{\beta}-1}{s_{\hat{\beta}}}$, with $s_{\hat{\beta}}^2 = \frac{n^{-1}\sum_{t=1}^n (y_t - \hat{\beta} y_{t-1})^2}{y_{t-1}^2}$, has a weak limit:

$$t_\beta \xrightarrow{w} \frac{\int_0^1 L_\alpha(s-) dL_\alpha(s)}{\sqrt{[L_\alpha](1) \int_0^1 L_\alpha^2(s) ds}}. \tag{6.20}$$

Adding a drift and keeping the prior assumptions for u_t, results in $\Delta y_t = \mu + \beta y_{t-1} + u_t$ (case II). The OLS estimator for β is then

$$\hat{\beta} = \frac{\sum_{t=1}^2 y_t y_{t-1} - 1/n(\sum_{t=1}^n y_{t-1})(\sum_{t=1}^n y_t)}{\sum_{t=1}^n y_{t-1}^2 - 1/n(\sum_{t=1}^n y_{t-1})^2} \tag{6.21}$$

Assume that the assumptions on postulations regarding $\hat{\beta}$ hold. Under the null hypothesis $(\mu = 0, \beta = 1)$, the OLS estimator for μ is

$$\hat{\mu} = \frac{\sum_{t=1}^n y_t \sum_{t=1}^n y_{t-1}^2 - \sum_{t=1}^n y_{t-1} \sum_{t=1}^n y_t y_{t-1}}{n \sum_{t=1}^n y_{t-1}^2 - (\sum_{t=1}^n y_{t-1})^2} \tag{6.22}$$

The t-statistics $t_{\hat{\beta},\hat{\mu}} = \frac{(\hat{\beta}-1)}{s_{\hat{\beta}}}$ with $s^2_{\hat{\beta}} = \frac{n^{-1}\sum_{t=1}^{n}(y_t-\hat{\mu}-\hat{\beta}y_{t-1})^2}{\sum_{t=1}^{n}y^2_{t-1}-n^{-1}\left(\sum_{t=1}^{n}y_{t-1}\right)^2}$, converges weakly as $n \to \infty$.

As mentioned, cointegration can be roughly interpreted as the situation when variables are trended but a linear combination between them is, in the long run, stationary. This defines a long-term equilibrium relationship. The representation theorem of Engle and Granger (1987) states that if there is a set of cointegrated variables then there exists a valid error correction representation of the data. For more detailed information on cointegration under the stable assumption, the reader is referred to Rachev and Mittnik (2000, chapt. 15).

6.3 Conclusion

This chapter has specified the idea of modeling corporate bond returns of different credit qualities with equal maturity as a system of equations by a cointegrated vector-autoregressive model.

There are two main reasons for this approach:

- A long-term relationship between zero-bond log price paths of different credit qualities can be observed for given time to maturity. The paths exhibit similar trends. The approach presented succeeds in capturing these effects.
- The problem of changing time to maturity that came from working with observed prices of real individual bonds can be resolved.

Initial empirical analysis demonstrates that the treasury returns are a significant risk driver for all corporate credit qualities (for equal time to maturity).

Selecting a set of defined maturities, a cointegrated VAR model consisting of one equation for each credit quality can be fitted for each of them.

Finally, the returns of a given credit quality can be obtained for each maturity by interpolating between the returns of the defined set of maturities.

To sum up, this chapter has presented a review of unit root theory and the order of integration, the theory of cointegration, and the concept of error correction models. Furthermore, unit root tests have been introduced. For unit root testing, the most common regression equations for the Dickey-Fuller test have been discussed.

As credit returns are known to follow a stable law, the last section has been dedicated to unit roots and cointegration in the stable case.

Chapter 7

Cointegrated VAR

This chapter reviews how to specify a vector error correction model. It presents methods for determining the three specification features:

- the lag order.
- the cointegration rank according to Johansen and an alternative method.
- the estimation of the cointegrating relations.

7.1 Lag Order Of The VECM

When building a cointegrated VAR model, the cointegration rank is typically not known at the beginning when the decision about the lag order has to be made. It is therefore sensible to focus on the unrestricted VAR first (Lütkepohl and Breitung, 1996). This form should be used to determine the lag order. Thus, the autoregressive order of the model is derived before the cointegration rank is known. There are a number of ways to determine the lag order. Among the most common are the Akaike Criterion and the Schwarz Criterion. Both are of the form

$$Cr(m) = log(det(\tilde{\Sigma}_u(m))) + c_T\phi(m), \tag{7.1}$$

with m as the tested lag length and T as the sample size. $\tilde{\Sigma}_u(m)$ is the residual covariance matrix estimator for a model of order m. $log(det(\tilde{\Sigma}_u(m)))$ measures the fit with the model. $c_T\phi(m)$ is a term imposing a penalty for an increasing number of lags. The test proposes the selection of that m which delivers the lowest value for the term $Cr(m)$. The exact specifications of the Akaike (AIC(m)) and Schwarz (SC(m)) criteria are given by[1]

[1]See Akaike (1973), Akaike (1974) and Schwarz (1978).

$$AIC(m) = log(det(\tilde{\Sigma}_k(m))) + 2(m)k^2/T, \tag{7.2}$$

and

$$SC(m) = log(det(\tilde{\Sigma}_k(m))) + (m)k^2 log(T)/T. \tag{7.3}$$

where k is the number of equations in the system and T is the sample size. The restrictions in a cointegrated VAR reduce the dimensionality of parameter space. Such restrictions are based either on economic theory or other assumptions.

Alternatively, t-ratios and F-tests are also common tools for testing the lag length. Unfortunately, problems may arise when they are applied to the VAR in levels before the cointegration rank is known although they keep their usual properties when checking for the short-run parameters of a VECM.

Moreover, sequential testing procedures can be applied:

$$\begin{aligned} H_0^{(1)} : p = p_{max} - 1(A_{p_{max}} = 0) \quad &\text{against} \quad H_1^{(1)} : p = p_{max}(A_{p_{max}} \neq 0); \\ H_0^{(2)} : p = p_{max} - 2(A_{p_{max}-1} = 0) \quad &\text{against} \quad H_1^{(2)} : p = p_{max} - 1(A_{p_{max}-1} \neq 0); \\ ...etc... \end{aligned}$$

Such tests are performed by applying a likelihood ratio or Wald test. It terminates when the null hypothesis is rejected for the first time.

7.2 Estimating Cointegrating Relations When Cointegration Rank is Known

The procedure introduced in this section is based on the assumption that there is currently one cointegrating relation among the variables of a VAR (Lütkepohl, 1994). The relationship is assumed to be linear with stochastic regressors. The VAR consists of k dependent variables, each with a time series of length t. The cointegrated variable is known. Thus, the variables of the VAR may be divided into two groups:

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}, \tag{7.4}$$

where y_{1t} consists of the cointegrated variable and y_{2t} is a $k-1$-dimensional vector containing the variables that are not cointegrated. The cointegrating relation is

$$\begin{aligned} y_{1t} &= \eta + y_{2t}'\gamma + u_{1t} \\ \Delta y_{2t} &= u_{2t}, \end{aligned} \tag{7.5}$$

where y_{1t} consists of a single variable only. u_t is stationary around mean zero and η and γ are parameters. Estimating η and γ, one obtains by least squares:

$$\begin{bmatrix} \hat{\eta} \\ \hat{\gamma} \end{bmatrix} = (Y_2'Y_2)^{-1}Y_2'y_1, \tag{7.6}$$

where

$$Y_2 = \begin{bmatrix} 1 & \dots & 1 \\ y_{2,1} & \dots & y_{2,T} \end{bmatrix} \text{ and } y_1 = \begin{bmatrix} y_{1,1} \\ . \\ . \\ . \\ y_{1,T} \end{bmatrix}. \tag{7.7}$$

Assuming the y_2 are strictly exogeneous and the residuals follow a normal distribution, then

$$u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \Sigma_{u2} \end{bmatrix} \right). \tag{7.8}$$

Then the least square estimators follow a normal distribution. For this reason, the usual test statistics and confidence intervals may be employed.

$\sqrt{T}(\hat{\eta} - \eta)$ and $T(\hat{\gamma} - \gamma)$ have a regular asymptotic distribution: because of T, $\hat{\gamma}$ has a faster convergence. It is therefore a *superconsistent* estimator.

The system above is called a *triangular system.*[2] The case above has a single cointegrated variable, but, of course, vector y_1 may consist of more than one cointegrated variable. A *triangular system* requires pre-information on the cointegrating relations: it is essential to know which of the dependent variables are cointegrated and which are not beforehand. There exists a cointegration matrix containing the cointegration vector for each cointegrated variable. r denotes the number of cointegrated variables. Thus, y_{1t} is an r-dimensional vector and y_{2t} is a $k-r$-dimensional vector. The variables in y_{2t} are not cointegrated.

For the case with $r < k$ cointegrated variables, the least squares estimator for γ is analogously given by

$$\hat{\gamma} = (Y_2'Y_2)^{-1}Y_2'Y_1\ . \tag{7.9}$$

[2]See Maddala and Kim (1998)

$\hat{\gamma}$ is superconsistent, however, its asymptotic distribution depends on nuisance parameters, as the regressors are endogeneous and serial correlation in the errors exist (Park and Philipps, 1988).

In order to overcome the problems related to endogeneity and serial correlation, a number of modifications have been presented in order to make the estimates more efficient. Both Saikkonen (1991) and Philipps and Loretan (1991) examined a model that included leads and lags of y_{2t} (Maddala and Kim, 1998).

Having the triangular system in error correction model (ECM) form, the pre-information about the r cointegrated variables is applied for estimation:

$$\Delta y_t = v + \alpha\beta y_{t-1} + \Gamma_1 \Delta y_{t-1} + ... + \Gamma_{p-1}\Delta y_{t-p+1} + \epsilon_t, \tag{7.10}$$

where $\hat{\alpha\beta}$ is obtained by a direct least squares approach. As it is known that $\beta = [I_r, -\beta_1]$, the first r columns of $\hat{\alpha\beta}$ may be used for $\hat{\alpha}$.

A least squares approach is then applied to the remaining $k - r$ columns of $\hat{\alpha\beta}$ denoted by H. β_1 is then obtained by

$$\hat{\beta}_1 = -(\hat{\alpha}'\hat{\Sigma}_\epsilon^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\hat{\Sigma}_\epsilon \hat{H}, \tag{7.11}$$

where $\hat{\Sigma}_\epsilon = \sum_{t=1}^{T} \hat{\epsilon}_t\hat{\epsilon}_t'/T$ is a usual estimate for the covariance matrix of the residuals. In general, the estimates are asymptotically efficient. The estimator for $\hat{\alpha\beta}$ is superconsistent in the single-equation case.

7.3 Estimating Cointegrating Relations When Cointegration Rank Is Unknown

When analyzing VAR-based cointegration with more than two variables, the number of cointegrating relations is not known in most cases. Thus the first step is to determine this parameter. This number is equal to the number of cointegrated variables.

Assuming there is a VAR of order p,

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + Bx_t + \epsilon_t \ , \tag{7.12}$$

where y_t is a vector of k non-stationary I(1) variables (variables in levels). x_t is a vector of $I(0)$ exogeneous variables, ϵ_t is the vector of innovations. The VAR can be written as an error correction model:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} ... + \Gamma_{p-1}\Delta y_{t-p+1} + Bx_t + \epsilon_t \ . \tag{7.13}$$

The matrix Π can be rewritten as[3]

$$\Pi = -(I_k - A_1 - A_2 - ... - A_p), \tag{7.14}$$

with I_k as the k-dimensional unity matrix.

The Γ_i are the negative sum of all matrices A_j for $j = i+1, ..., p-1, p$:

$$\Gamma_i = -(A_{i+1} + A_{i+2} + ... + A_p),\ i = 1, ..., p-1, \tag{7.15}$$

$$\Gamma_{p-1} = -A_p. \tag{7.16}$$

Equation (7.13) is the basic reduced-form error correction model for cointegration in terms of a reduced-rank hypothesis on matrix Π, as presented in Johansen (1995). Π defines the cointegrating vectors β and the adjustment coefficients α.

Recalling the two-variable case with y_{1t} and y_{2t} and the cointegration equation $y_t + \beta' x_t + u_t$ and the vector error correction (VEC) mechanism $\Delta y_t = \alpha_1 \Delta x_t + \alpha_2(y_{t-1} - \beta x_{t-1}) + \epsilon_t$, this can be expressed in vector notation as

$$\Delta y_t = \alpha_1 \Delta x_t + \alpha_2(1, -\beta)(y_{t-1}, x_{t-1})' + \epsilon_t. \tag{7.17}$$

If Δy_t and Δx_t are stationary, β is replaced by its OLS estimate $\hat{\beta}$.

Focusing on the cointegrated VAR model (7.13) with $k > 2$ variables, Π is the product of the matrices α and β. Its coefficients reflect the impact of the one-period lagged y_{t-1} on Δy_t. Both α and β are $k \times r$ matrices:

$$\alpha\beta' = \Pi \tag{7.18}$$

where matrix α contains the adjustment parameters of the VEC model. One parameter of each row in matrix β is equal to 1. A row represents one cointegrating vector. The rank of Π indicates the number of cointegrating relations, which is the cointegration rank. It is the maximum number of linearly independent columns in the matrix. For the rank of a product of two matrices, the following relationship holds: $r(AB) \leq min(r(A), r(B))$. Thus, the rank of Π lies between zero and k.

The above statements are generalized by the so-called *Granger representation theorem*:[4]

1. If the rank of the matrix Π is equal to the number of variables k in the VAR, then the vector process y_t is stationary, i.e. all the variables in y_t are integrated of order zero, I(0).

[3] See Johansen (1991) and Johansen (1995).
[4] See Charemza and Deadman (1997).

2. If the rank of Π is $r < k$, then there exists a representation of Π such that $\Pi = \alpha\beta'$. α, β are $k \times r$ matrices. β is called the cointegrating matrix and it has the following properties:
   - βy_t is I(0) and y_t is I(1).
   - The cointegrating vectors $\beta_1, \beta_2, ..., \beta_r$ are particular rows of the cointegrating matrix β.

The cointegration rank of a VAR can be determined by the Johansen procedure. This procedure is introduced later in detail. It considers the above reduced-form error correction model of the form

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-1} + Bx_t + \epsilon_t \tag{7.19}$$

under the reduced rank condition

$$\Pi = \alpha\beta'. \tag{7.20}$$

If the VAR has k endogeneous variables, each having a unit root, there could be from 0 to $k-1$ cointegrating relations. For each cointegrating relation, an additional error correction term of the form $\alpha_{ij}\beta_j' y_t$ is added to each equation of the VAR, where α_{ij} is a constant and β_j is a k-dimensional cointegrating vector.

If there are k cointegrating relations, none of the series has a unit root. The VAR may be represented solely by the levels y_t of the time series involved. This refers to the I(1) model, which is named as $H(k)$. In the reduced-form representation above, the levels are only lagged in the first differences. The following expresses the relations among the possible restrictions for the VAR. The formulation shows that the I(1) models form a nested sequence of models.

$$H(0) \subset ... \subset H(r) \subset ... \subset H(k). \tag{7.21}$$

$H(0)$ corresponds to the restriction $\Pi = 0$, which is the VAR model based on the process in differences.

7.4 Determining The Cointegration Rank Of A VAR

As mentioned above, Johansen presented a system estimation method in order to identify the rank of a cointegrated VAR and to determine the parameters of its reduced-form error correction model (see equation (7.13)). The procedure is based on a likelihood ratio test with Johansen assuming that the error terms follow a Gaussian law. Johansen has shown that the first r estimated eigenvectors are the maximum likelihood (ML) estimates of the system's cointegrating vectors. The procedure performs a hypothesis testing, starting with the testing of the null hypothesis $H_0^0 : r = 0$ (no cointegrating relations) versus $H_1^0 : r = 1$ first, and then performing tests of

$$\begin{gathered} H_0^1 : r = 1 \text{ versus } H_1^1 : r > 1, \\ H_0^2 : r = 2 \text{ versus } H_1^2 : r > 2, \\ ...., \\ H_0^k : r = k-1 \text{ versus } H_1^k : r = k. \end{gathered}$$

This test is called the *trace test*. It checks the hypothesis that there are at most r cointegrating vectors.

Another testing procedure, called *maximum eigenvalue test*, examines the hypothesis $H_0 : r$ cointegrating vectors versus $H_1 : r+1$ cointegrating vectors.

Both the trace test and the maximum eigenvalue test terminate the first time the null hypothesis is no longer rejected.

Focusing on the VAR in the known VECM form

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + ... + \Gamma_{p-1} y_{t-p+1} + \epsilon_t, \tag{7.22}$$

for the Johansen test, in order to be consistent, $\alpha\beta'$ should not be full rank since $\Delta y_{t-1} ... \Delta y_{t-p+1}$ are all I(0) but y_{t-1} is I(1).

Let the rank be r. To execute the Johansen test, the Γ_i are eliminated first by regressing Δy_t on $\Delta y_{t-1} ... \Delta y_{t-p+1}$. R_{0t} are the residuals of the regression. Moreover, y_{t-1} is regressed on $\Delta y_{t-1} ... \Delta y_{t-p+1}$ with R_{1t} as the residuals. Thus, the regression equation can be reduced to

$$R_{0t} = \alpha\beta' R_{1t} + u_t \ . \tag{7.23}$$

The regression problem is multivariate. The sums of the squared residuals are denoted as follows:

$$S_{00} = \sum_{t=1}^{T} R_{0t}^2, \text{ and} \tag{7.24}$$

$$S_{11} = \sum_{t=1}^{T} R_{1t}^2, \tag{7.25}$$

where T is the size of the sample.

Moreover,

$$S_{10} = S_{01} = \sum_{t=1}^{n} R_{0t} R_{1t} \tag{7.26}$$

provides the matrix $\begin{bmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{bmatrix}$ of sums of squares and sums of products for R_{0t} and R_{1t}. Each of these matrices is of order $k \times k$.

First, the likelihood function is maximized with respect to α, holding β constant. In the second step, the maximization with respect to β is performed. $\hat{\alpha}'$ is an $r \times k$ matrix and the conditional maximum likelihood function is given by

$$\hat{\alpha}' = (\beta' S_{11} \beta)^{-1} \beta' S_{10} \; . \tag{7.27}$$

The conditional likelihood function for β is then

$$[L(\beta)]^{-2/T} = |S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta S_{10}|. \tag{7.28}$$

Maximization of the likelihood means that the determinant on the right of the equation has to be minimized. The determinant can be reformulated as follows:

$$\frac{|\beta' S_{11} \beta' S_{10} S_{00}^{-1} S_{01} \beta'| \cdot |S_{00}|}{|\beta S_{11} \beta|} \tag{7.29}$$

The elements of the matrix α determine the speed of adjustment in terms of a disturbance. After normalization, the elements of matrix β are interpreted as long-run parameters.

Thus, the maximum of the likelihood function is given by minimizing the upper equation. This is performed, after some substitutions and reformulations, by solving the following eigenvalue problem:

$$|S_{10} S_{00}^{-1} S_{01} - \lambda S_{11}| = 0 \text{ , or finding the eigenvalue of} \tag{7.30}$$

$$|S_{11}^{-1} S_{10} S_{00}^{-1} S_{01} - \lambda I| = 0. \tag{7.31}$$

If λ_i are the canonical correlations given by solving equation (7.31), then $(1-\lambda_i)$ are the eigenvalues of

$$(I - S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}). \tag{7.32}$$

Since the value of the determinant of a matrix is equal to the product of its eigenvalues, one obtains

$$\prod_{i=1}^{k}(1-\lambda_i) = |I - S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}|. \qquad (7.33)$$

The maximum of the likelihood is finally given by

$$L_{max}^{-2/T} = |S_{00}|\prod_{i=1}^{k}(1-\lambda_i). \qquad (7.34)$$

The procedure has just been briefly sketched here. For a more detailed presentation, please see Maddala and Kim (1998).

7.5 The Trace Test And The Maximum Eigenvalue Test

The Johansen procedure leads to two test statistics for the rank of cointegration. The first is called the *trace test*. It tests the hypothesis that there are, at most, r cointegrating vectors.

The likelihood ratio test statistic for at most r cointegrating vectors (trace test) is

$$\lambda_{trace} = -T\sum_{i=r+1}^{k} log(1-\hat{\lambda}_i), \qquad (7.35)$$

where $\hat{\lambda}_{r+1}, ..., \hat{\lambda}_k$ are the $(k-r)$ smallest eigenvalues of the determinant equation (7.31).

The maximum eigenvalue test examines the null hypothesis of r cointegrating vectors versus the hypothesis of $r+1$ cointegrating vectors. The likelihood ratio test statistics is

$$\lambda_{max} = -Tlog(1-\hat{\lambda}_{r+1}). \qquad (7.36)$$

Both tests are based on eigenvalue-eigenvector decomposition of matrix $\alpha\beta'$. The critical values (quantiles expressing the confidence levels) are tabulated. These values increase with the size of the VAR system. The tables with the corresponding values are given in detail in Osterwald-Lenum (1992).

The Johansen test was extended to include trends and drifts. Three cases are distinguished:[5]

1. Error correction models without a drift.

[5] Maddala and Kim (1998).

2. Error correction models with separate drifts.
3. Error correction models with constants in their error correction.

The trace test or maximum eigenvalue test is performed for an ascending number of cointegrating relations. Testing is carried out until the null hypothesis is no longer rejected for the first time.

McKinnon (1991) calculated the approximated values for both tests considering up to 12 cointegrating relations.

7.6 Determining Cointegration Rank With Model Selection Criteria

Most popular in determination of cointegration rank are the sequential testing procedures with ML estimates proposed by Johansen. As presented, the two possible testing sequences are maximum eigenvalue and trace test. Alternatively, the model selection criteria can be applied to consistently estimate the rank.

Thus, the concept of Akaike Criterion and Schwarz Criterion may be set up for the determination of the cointegration rank of a system as well. The statistic for rank determination is also of the form

$$C_T(r) = Q_T(r) + \rho(r)c_T \text{ with } r = 0...k. \tag{7.37}$$

where $\rho(r)$ is any strictly monotonically increasing function of r and c_T is a sequence of numbers which is $o(T)$. $Q_T(n)$ is defined to be zero. n is the number of variables in the system. For the cointegration rank the r is chosen that minimizes $C_T(r)$.

Suppose, $Q_T(r_0)$ is a test statistic for testing the null hypothesis $H_0(r_0)$, with the properties

$$\begin{aligned} plim\ Q_T(r)/T > 0 \text{ for } r < r_0 \text{ and} \\ Q_T(r) = O_p(1) \text{ if } r \geq r_0. \end{aligned} \tag{7.38}$$

For the application of the Akaike Criterion and Schwarz Criterion in this context, the reader is referred to Lütkepohl (1998).

A further option for specification of the cointegration rank of a system is the Stock-Watson test (see Lütkepohl, 1998; Stock and Watson, 1988).

If the rank of cointegration has been determined, the coefficients of the VECM can be obtained by least squares estimation. In case of lag order 1, the estimate for Γ is then

$$\hat{\Gamma} = (\Delta Y - \Pi Y_{-1})X'(XX')^{-1} \tag{7.39}$$

With $M = I - X'(XX')X$, the substitution $\Delta YM = \Pi Y_{-1}M + \hat{U}$ can be performed and $\hat{\Pi}$ estimated. However, as $\Pi = \alpha\beta'$, both $\hat{\alpha}$ and $\hat{\beta}$ have to be determined separately. β is obtained according to Johansen (1995):

- First define $S_{00} = n^{-1}\Delta YM\Delta Y'$, $S_{01} = n^{-1}\Delta YM\Delta Y'_{-1}$, and $S_{11} = n^{-1}\Delta Y_{-1}M\Delta Y'_{-1}$. n is the number of observations.

- Then, the eigenvalue problem $det(\lambda S_{11} - S'_{01}S_{00}^{-1}S_{01}) = 0$ has to be solved. The eigenvalues are ordered according to $\lambda_1 \geq ... \geq \lambda_k$ where v_i are the corresponding eigenvectors. The eigenvectors satisfy $\lambda_i s_{ii} v_i = S'_{01}S_{00}^{-1}S_{01}$. The normalized eigenvectors which satisfy $V's_{11}V = I_k$ are the estimates of the cointegrating vectors $\hat{\beta}$.

- Having $\hat{\beta}$, $\hat{\alpha}$ is obtained by OLS estimation

$$\hat{\alpha} = \Delta YMY'_{-1}\hat{\beta}(\hat{\beta}'Y_{-1}MY'_{-1}\hat{\beta})^{-1}. \tag{7.40}$$

- Thus, the estimator of Π is then $\hat{\Pi} = \hat{\alpha}\hat{\beta}'$ Under the Gaussian assumption, the OLS estimates refer to the ML estimates conditional on the presample values (Johansen, 1988).The estimates of Π and Γ are consistent and asymptotically normal under general assumptions.

It is important to mention that the obtained $\hat{\beta}$ might not refer to the econometric identification. $\hat{\beta}$ solely describes the cointegration space which is estimated with this procedure. It allows several possible linear combinations of cointegrating vectors. Thus, other identifying restrictions for the product $\alpha\beta' = \Pi$ need to be imposed. In order to make β unique, so-called uniqueness restrictions are applied. In this case, β is specified to have a left part that is an $r \times k$-dimensional identity matrix with r as the cointegration rank and k the number of endogeneous variables, yielding $[I_r\beta'_1]$. If $r = 1$, this leads to normalizing the coefficient of the first variable. With these uniqueness restrictions, $T(\hat{\beta} - \beta)$ and $\sqrt{T}(\hat{\alpha} - \alpha)$ converge in distribution (Johansen, 1995). As $\hat{\beta}$ converges faster with rate T, the estimator is called superconsistent. The covariances of the cointegrated VAR are obtained in the same way as for the unrestricted VAR.

7.7 Cointegration In Credit Modeling

The cointegration tests help in solving the questions of whether a VAR system should be modeled in levels or in differences - or both with some restrictions. The cointegrating relations define a *long-term relationship* among the endogeneous variables involved.

In the given case, the concept of cointegration will be applied to model the relationship among US market daily credit returns for each credit rating and each maturity. Not only are the returns for different credit ratings driven by a common exogeneous factor, the treasury rate, they also show a common behavior which cannot be explained by exogeneous factors.

The historical daily credit returns are obtained from daily yield curves for given credit ratings. They represent the average yield to maturity for the given creditworthiness. Such average yield curves are calculated from the daily prices of numerous traded corporate bonds with equal credit rating and equal time to maturity.

While series of daily log prices of bonds are not stationary, the differences (returns) exhibit stationary behavior. The objective is to construct a cointegrated VAR model that describes the behavior of the bond returns, yet also considering the long-term behavior of the variables in levels (log prices).

Johansen's test focused on models where all the variables depend on each other and none of them is exogeneous. This is considered as the closed form. But, when describing corporate bond returns, the development of the treasury bond's price is the major driver of the system.

Of course, in the given case it would have been possible to model the price of the treasury bond as another dependent variable of the system. The decision not to do so has two main reasons:

- First, consistency with the initial single-equation model for individual bonds - presented in sections 3.7 and 3.6 - is assured. Those models have the log returns of the riskless asset as an exogeneous factor.

- Second, the cointegrated VAR model would become more complex by choosing seven dependent variables.

- And, it is implied by economic understanding, that a change in the treasury yield curve causes an effect on corporate bond returns. Yet there is not necessarily an influence of the corporate bond returns on the treasury returns. Corporate bond returns hinge on several factors aside from the treasury returns - e.g. market liquidity, average credit risk, etc. Thus, the treasury returns cause changes in corporate bond returns but the treasury returns themselves are not understood to be affected by representations of the corporate bond prices or returns.

7.8 Conclusion

This chapter has reviewed possible methods to determine the specification of the cointegrated VAR (lag order, cointegration rank) and to estimate the parameters. It has dealt both with the case where the cointegrated variables are known in advance and with the case where the cointegrating vectors are to be determined. For future modeling under the stable assumption, applying the Johansen procedure based on the Gaussian assumption even for the stable case is suggested. To do so, the procedure is then applied to the truncated time series sample. In the next chapter, a closer look will be taken at exactly this issue.

Chapter 8

VAR Models For Credit Returns

In this chapter, three stable cointegrated VAR models based on the available data are specified and built. Each VAR describes the corporate bond returns of six rating grades for a given time to maturity. The chosen maturities are 2 years, 10 years and 30 years. Section 8.5 analyzes the behavior of the treasury returns and develops an appropriate model for their description.

The building of the cointegrated VAR models proceeds in the following manner:

- The preconditions for cointegration (I(1) in levels and I(0) in differences) are checked by unit-root tests, applying an appropriate testing strategy.

- The lag order of the unrestricted VAR in levels is determined by the Schwarz Criterion and the Akaike Criterion.

- The "traditional" cointegrating relations and cointegration rank are specified with the Johansen procedure. However, as the innovations of the VAR are assumed to follow a stable law, further considerations and adjustments have to be made before applying the procedure.

The main findings of this chapter are:

- The "traditional" cointegrating relations based on the Johansen procedure are found too weak to describe the real long-term behavior of the price paths. Thus, an alternative set of cointegrating relations is developed that makes the intersecting of neighboring price paths less likely.

- Besides the treasury returns there is another common risk factor driving the returns of all corporate rating grades: This risk factor represents the common changes of credit spread over all rating grades.

- The influence of the treasury returns on the corporate bond returns is the greater the better the credit quality of the corporate bonds. The daily log prices of the treasury bonds are found to follow a mean-reverting process.

- The residuals of the cointegrated VAR models exhibit signs of volatility clustering. This phenomenon will be treated in chapter 9.

8.1 The Data And Variables

The model introduced in this chapter is capable of describing the returns of corporate zero-bonds depending on credit risk (represented by rating grades) and maturity. As observed in the chapter 6, for a given maturity the average log prices of different corporate credit ratings seem to have common long-term trends. In addition, they are driven by the returns of treasury bonds with equal maturity as an exogeneous variable.

According to Park (1997) it is generally accepted in the literature that treasury bill yields are not stationary and follow an I(1) process.[1] It is reasonable that this can be assumed for the log prices of corporate zero-bonds with fixed time to maturity as well. Tests will prove this later.

The time series of log prices for the corporate zero-bond were derived from daily yield-to-maturity data. The historical fair market yield curves for different corporate bond ratings and different maturities were obtained from the Bloomberg System. The returns are obtained as the difference of the log prices of two subsequent days. In addition, the returns of treasury bonds with equal maturity were calculated as well.

For the chosen time period of 8/1996 through 7/2000, daily yield-to-maturity data of corporate bonds for six credit rating grades and various maturities were available within the US bond market. These yields represent averages for a given rating category and maturity, derived from the prices of numerous traded US corporate bonds of the industrial sector. The yields are presented for the corporate credit rating grades AAA, AA, A, BBB, BB, B.[2] Data has been chosen for three maturities: 2 years, 10 years, and 30 years. The sample size of each time series is 1044.

For the purpose of VaR, the interest is on the average daily price movement for bonds of a given credit rating i and given time to maturity T. Therefore, the variables explained by the model are the credit returns at time t. Generally, at time t the return $R_{t,T}$ of a bond with maturity T is defined the following way:

[1] Park, for example, has tested this for Canadian treasury bill yields.

[2] Standard & Poors rating grade system.

$$R_{t,T} = log(P_{t,T}) - log(P_{t-1,T-1}) \ , \tag{8.1}$$

where $P_{t,T}$ is the price of the bond with maturity T, valued at time t. The prices are driven by the daily movements of the corporate bond yield curves. Note that for the definition of bond returns in equation (8.1) the maturity $T - t$ in the numerator is identical to the maturity $T - 1 - (t - 1)$ in the denominator.

The average log prices of corporate zero bonds with Standard & Poors credit grades AAA, AA, A, BBB, BB, B are plotted over the time period from 8/96 to 8/00, having one chart for each maturity. It is not surprising that they clearly show patterns of cointegration as the paths exhibit equal trends. However, the trends do not look deterministic.

The charts of log prices for the three maturities are presented in figures 8.1 - 8.3.

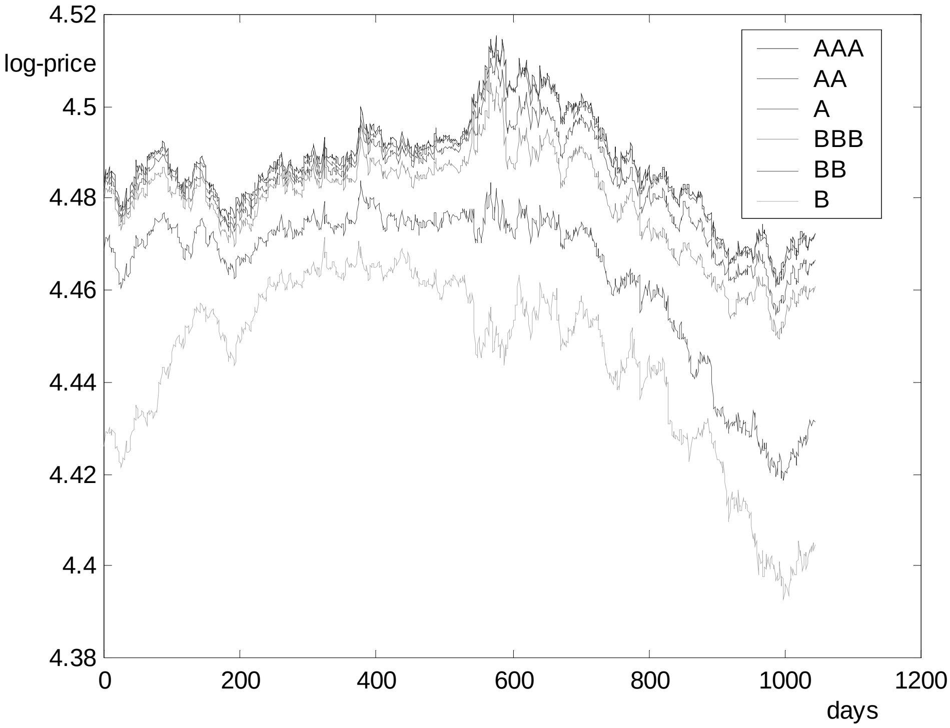

Figure 8.1: Log prices of 2-year maturity corporate zero bonds.

As argued in chapter 6, a suitable approach for capturing such a long-term behavior with a system of equations is the application of a *cointegrated vector-*

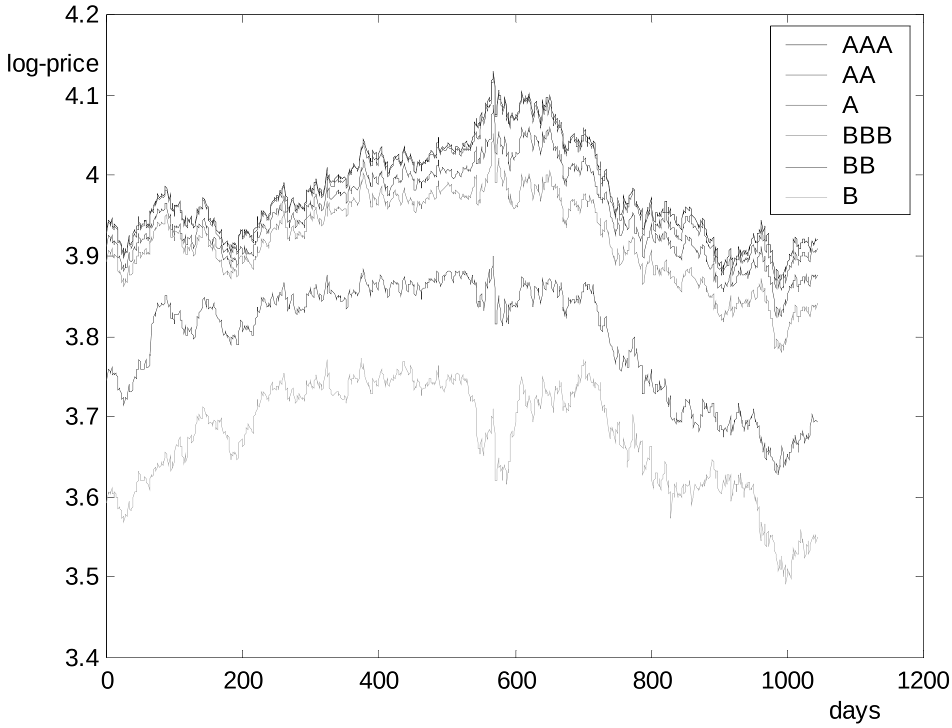

Figure 8.2: Log prices of 10-year maturity corporate zero bonds.

autoregressive (VAR) system. Here the system for each maturity has six equations, one for each rating grade. $log(P_{t,T})$ are the levels that should be integrated of order 1, and $R_{t,T}$ are the differences that should be integrated of order 0. This will be examined by a unit root test first. Furthermore, the unit root test is also applied to check the cointegrating relations.

8.2 Testing For Unit Roots

The testing for unit roots should consider prior knowledge on the observed time series in order to exclude unrealistic outcomes and make the testing less complicated. Following Elder and Kennedy (2001), such a testing strategy will be applied here. By looking at the data of corporate bond log prices, it can be easily seen that those are not growing. It makes sense to exploit such information before selecting the test.

The general testing of the augmented Dickey-Fuller (ADF) is performed with the equation

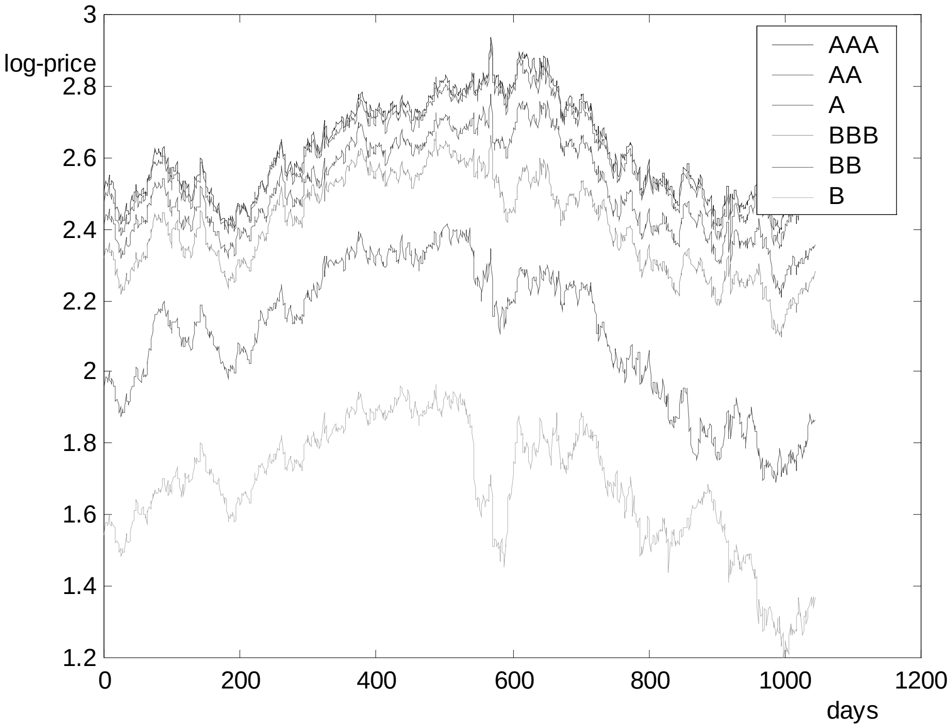

Figure 8.3: Log prices of 30-year maturity corporate zero bonds.

$$\Delta y_t = (\rho - 1)y_{t-1} + \alpha + \beta t + \epsilon_t, \tag{8.2}$$

and it is tested for $\rho = 1$ which indicates the presence of a unit root.

In the case of log prices for corporate zero-bonds the presence of a time trend β can be denied. Analyzing the resulting equation,

$$\Delta y_t = (\rho - 1)y_{t-1} + \alpha + \epsilon_t, \tag{8.3}$$

there would be only a unit root if there was a zero intercept. In this case, the joint hypothesis $\rho = 1$ and $\alpha = 0$ has to be tested. If this null hypothesis is rejected, it can be concluded that the process is nonstationary and has a mean zero. This could be shown by applying an F-test. However, if $\rho = 1$, it is not reasonable if α was not zero. Therefore a t-test would be preferable as it has a higher power because it is one-sided. For the null-hypothesis, however, the t-statistic does not have a t-distribution. The special critical values can be obtained by Monte Carlo studies. They are available from many text books (e.g. Hamilton, 1994).

The unit root tests for the log prices $y_{i,t}$ were performed with an ADF-test routine in Matlab. For all rating grades and maturities, the absolute values for the test statistic were less than the absolute value of -2.594, which is the critical value for the 10% level. It can be concluded that the log prices are I(1), i.e. integrated of order 1. Contrary, the corresponding log price differences $\Delta y_{i,t}$ are found to be stationary, and thus are I(0).

8.3 Specification Of The VECM

The dependent variables on the left side are the corporate zero-bond returns for a fixed time to maturity. The returns of a treasury zero-bond with equal maturity enter the VAR as an exogeneous variable. The resulting Cointegrated VAR is represented by a Vector Error Correction Model (VECM) (Charemza and Deadman, 1997), with

$$\Delta y_{t,T} = \Pi y_{t-1,T-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j,T-j} + B x_{t,T} + res_t, \tag{8.4}$$

where p is the order of the unrestricted VAR, $\Delta y_{t,T}$ is the vector of returns, $y_{t-1,T-1}$ is the vector of lagged log prices, and $x_{t,T}$ are the treasury returns. The matrix Π can be decomposed as $\Pi = \alpha\beta'$. The innovations $res_{t,i}$, $i = 1 \ldots 6$, for AAA, AA, A, BBB, BB and B, are a six-dimensional symmetrically stable random vector, with $1 < \alpha < 2$. For the properties of stable non-Gaussian distributions, the reader is referred to section 2.2 and Samorodnitsky and Taqqu (1994, chapt. 1). The time series of the innovations are Lévy processes with symmetrically stable increments.

The VECM in (8.4) is modeled without a constant term as the dependent variables of the VECM have means around zero.

$$y_{t,T} = \begin{bmatrix} log(P_{t,T,AAA}) \\ log(P_{t,T,AA}) \\ log(P_{t,T,A}) \\ log(P_{t,T,BBB}) \\ log(P_{t,T,BB}) \\ log(P_{t,T,B}) \end{bmatrix} \tag{8.5}$$

The VECM in (8.4) represents one time to maturity $T-t$. Thus, for each of the maturities 2 years, 10 years and 30 years, a cointegrated VAR with six equations is built.

First, the unrestricted VAR is built as the lag length p is determined from the unrestricted VAR in levels. This is done before the cointegration rank is known. The test statistics applied in this case are the Akaike Criterion and the Schwarz Criterion. The lag order is obtained by minimizing these criteria for the unrestricted VAR. While the Akaike Criterion finds a lag order of $p = 2$, the Schwarz Criterion delivers $p = 1$.

The decision for the final choice of p between Akaike and Schwarz is made under the premise to select the one that keeps more information in the given case. Thus, p is set to be 2 (this results in a lag of $p - 1 = 1$ for the VECM). Therefore, the mean equation is then modeled as a cointegrated VAR of order 1.

Usually, cointegrated VARs do not have exogeneous variables. However, these may be added. As stated above, the treasury returns have a considerable influence on the movements of corporate bond prices. Therefore, the returns of a treasury zero bond with the same maturity as the endogeneous corporate bond returns will enter the right side of each equation in the form of an exogeneous variable. As known from the initial examinations in chapter 6, this is expected to explain a large fraction of the movements of the corporate bonds' prices. Performing a regression, the following results: *The higher the corporate credit rating, the greater is the influence of the treasury returns on a corporate bond's returns.*

After the determination of the lag order, the Johansen test is performed in order to obtain the cointegration rank and the cointegrating space.

As indicated in the previous chapter, the Johansen procedure has been developed under the Gaussian assumption.

The risk factors of the cointegrated VAR to be developed here are considered to follow a stable law. *However, the stable model of Johansen's test is untractable at the moment.* The literature does not provide rigorous results on stable Johansen as this has turned out to be an extremely difficult problem. Thus, this work tackles the problem as follows:

For the given case, the fact that the Johansen model is still valid when the innovations are in the Gaussian domain of attraction is used. To demonstrate this, the pre-limit theorems (Rachev and Mittnik, 2000) are applied.

The truncated stable distribution provides a superior fit of the innovations' distributions compared to the Gaussian fitting. On the other hand, the innovations are in the Gaussian domain of attraction. Therefore, the limiting results for the Johansen test are still valid.

Both the trace test and the maximum eigenvalue test provide a cointegration rank of 2.[3] A cointegration rank $r = 2$ means that the variables $log(P_{t,T,AAA})$

[3] The test was performed with a routine implemented in Matlab.

and $log(P_{t,T,AA})$ are cointegrated since the highest eigenvalues are obtained for them. These variables are explained by linear combinations of the other variables $log(P_{t,T,A})$, $log(P_{t,T,BBB})$, $log(P_{t,T,BB})$, and $log(P_{t,T,B})$.

The cointegrating relations have been tested both with and without a constant term. However, fitting the cointegrating relations with a constant turned out not to be significant. The normalized eigenvectors obtained for the cointegrating relations in case of $r = 2$ are

$$\beta_2 = \begin{bmatrix} 1 & 0 & -1.645 & 0.585 & 0.046 & 0.015 \\ 0 & 1 & -1.712 & 0.656 & 0.036 & 0.020 \end{bmatrix} \tag{8.6}$$

for the 2-year maturity cointegrated VAR,

$$\beta_{10} = \begin{bmatrix} 1 & 0 & -1.539 & 0.462 & -0.036 & 0.115 \\ 0 & 1 & -1.496 & 0.502 & -0.141 & 0.137 \end{bmatrix} \tag{8.7}$$

for the 10-year maturity cointegrated VAR, and

$$\beta_{30} = \begin{bmatrix} 1 & 0 & -0.973 & -0.285 & -0.039 & 0.348 \\ 0 & 1 & -1.035 & -0.116 & -0.093 & 0.281 \end{bmatrix} \tag{8.8}$$

for the 30-year maturity cointegrated VAR.

8.4 Revised Cointegrating Relations

The cointegrating relations obtained by the Johansen procedure neglect the fact that there is a certain order among the price paths. Thus, the chosen cointegrating relations should make it very unlikely that the price paths of different credit risk grades will intersect. The traditional cointegrating relations try to explain the relationship between the different paths by the lowest possible cointegration rank. Unfortunately, they permit crossing and changed order of the price paths. In order to reduce the likelihood of a changed order, the following restrictions are introduced:

$$y_{i,t} - y_{i+1,t} = c_i + u_{i,t} \ , \tag{8.9}$$

where $u_{i,t}$ is five-dimensional symmetrically stable, with $i = 1 \ldots 5$ for the first five credit rating grades. The estimated model is $y_i - y_{i+1} - c_i = 0$. The restrictions insure that the intersecting of neighboring paths becomes less likely. Now, there are five restrictions compared to the two restrictions obtained by Johansen's reduced rank regression. It should be noted that these relationships are different from the traditional cointegrating relationships, and have a cointegration rank of 5.

The plot of the log prices show that the distance between two neighboring price paths is relatively constant over time and varies generally around a mean. This behavior is described by the relations in (8.9). Fitting the cointegrating relations, the results for the β are given in 8.10 - 8.12.

$$\beta_2 = \begin{bmatrix} 1.0000 & 0 & 0 & 0 & 0 \\ -1.0000 & 1.0000 & 0 & 0 & 0 \\ 0 & -1.0000 & 1.0000 & 0 & 0 \\ 0 & 0 & -1.0000 & 1.0000 & 0 \\ 0 & 0 & 0 & -1.0000 & 1.0000 \\ 0 & 0 & 0 & 0 & -1.0000 \\ 0.0010 & 0.0031 & 0.0045 & 0.0152 & 0.0190 \end{bmatrix} \tag{8.10}$$

for the 2-year maturity cointegrated VAR and

$$\beta_{10} = \begin{bmatrix} 1.0000 & 0 & 0 & 0 & 0 \\ -1.0000 & 1.0000 & 0 & 0 & 0 \\ 0 & -1.0000 & 1.0000 & 0 & 0 \\ 0 & 0 & -1.0000 & 1.0000 & 0 \\ 0 & 0 & 0 & -1.0000 & 1.0000 \\ 0 & 0 & 0 & 0 & -1.0000 \\ 0.0055 & 0.0243 & 0.0295 & 0.1223 & 0.1242 \end{bmatrix} \tag{8.11}$$

for the 10-year maturity cointegrated VAR and

$$\beta_{30} = \begin{bmatrix} 1.0000 & 0 & 0 & 0 & 0 \\ -1.0000 & 1.0000 & 0 & 0 & 0 \\ 0 & -1.0000 & 1.0000 & 0 & 0 \\ 0 & 0 & -1.0000 & 1.0000 & 0 \\ 0 & 0 & 0 & -1.0000 & 1.0000 \\ 0 & 0 & 0 & 0 & -1.0000 \\ 0.0233 & 0.0873 & 0.0996 & 0.3049 & 0.4227 \end{bmatrix} \tag{8.12}$$

Co-integrating ADF test. For most of the cointegrating relations $y_{i,t} - y_{i+1,t} = c_i + u_{i,t}$, $i = 1 \ldots 5$, testing with the co-integrating ADF test clearly indicates that they are stationary and have the I(0) property. Thus, they are valid cointegrating relations. The absolute values of the t-statistic are greater than the absolute value of −3.038, which is the 10% critical value. The cointegrating relations are stationary and do not have a unit root, i.e. they are I(0). Thus, they are valid cointegrating relations.

However, for the cointegrating relation between the neighboring investment and speculative rating grades BBB and BB, the unit root hypothesis cannot be rejected at the 10% level. The same is true for the cointegrating relation between the two speculative grades BB and B. It can be concluded that cointegrating relations are weak not only between the investment grade and speculative grade area but also between the two speculative grades. However, there exist mechanisms that prevent the average price paths of neighboring rating grades from intersecting. Furthermore, it can be observed that the better the credit quality of neighboring grades, the stronger the cointegrating relations.

The restrictions in a cointegrated VAR reduce the dimensionality of parameter space. They come either from economic theory or from other constraints. However, the parameter space itself has no economic meaning. And while there are many linear combinations of cointegration vectors that can form the parameter space, economic meaning can be attributed only to certain cointegrating vectors.

The treasury returns $x_{t,T}$ are present as a risk driver in all equations of the VAR. For the AAA bonds, it explains about 73% of the price movements. For the B-rated bonds, it explains roughly 40%.

After estimating the VECM with OLSE, the residuals of the model show three attributes:

- Heavy-tailedness.
- A relatively high degree of correlation.
- Clustering volatility.

The first two properties seem familiar. Applying a stable fit to the residuals of the VECM, the dependence within the vector of the stable innovations can be captured via a sub-Gaussian vector. The residuals of the credit return model have a symmetrical distribution and can be fitted by $S\alpha S$ distributions. The $S\alpha S$ random vector can be represented by a product of a Gaussian random vector with dependent elements and a totally skewed $\alpha/2$-stable random vector which is independent of the Gaussian random vector.[4] By so doing, the dependence structure of the Gaussian vector can be transferred into the stable vector.

However, since the residual plots also show signs of clustering volatility, they imply a further examination of their behavior.

[4] Rachev, Schwartz and Khindanova (2001).

8.4.1 Checking Model Settings and Further Evaluation

Alternatively, the appropriate model specification for each single equation of the VAR could, for example, be obtained by applying the Box-Jenkins methods.[5] This provides the number of lagged variables or the specific mechanism that drives the variance (variation) of the error terms (for example, the resulting process could be a VAR-GARCH with an error correction mechanism ECM and with stable innovations).

With the Box-Jenkins method, the lags of an AR process are usually determined by analyzing the correlogram of the observed variable. Aside from the determination of the autoregressive lag, it has to be tested whether lagged I(0) variables of the other system variables have a significant impact as well. This can be performed by checking for Granger causality.

Once a model has been specified and estimated, there are numerous tests and procedures which may follow. Many tests are based on the residuals of the final cointegrated model (full residual vectors or single equations). However, the application of the coefficient of determination (R^2) as the traditional criterion for evaluation of the performance of an econometric model is not always suitable for the special characteristics of a VAR. Values of R^2 belonging to different models with different explanatory variables cannot be compared. A better criterion to assess a VAR is the coefficient of determination *adjusted for the number of explanatory variables.* Both t-values and F-statistic can be used to analyze the coefficients. Well-known are the Akaike Information Criterion and and the Schwarz Bayesian Criterion. Others are the final prediction error (FPE), the forecast Chi, the Durbin Watson statistic, or the Chow test.

As the cointegrated credit return model is supposed to work for Value at Risk (VaR) measurement, test statistics which evaluate the accuracy of the VaR forecast have to be applied. These will be introduced in a later chapter.

8.4.2 Analysis Of The Residuals

The residuals of each equation are plotted in a chart (see figure 8.4). The three properties mentioned above are analyzed in more detail now:

1. The residuals show a high degree of dependence and are therefore *highly correlated.*

2. In addition, the residuals have *peaked and heavy tailed distributions.* This supports the application of stable distributions with $\alpha < 2$.

[5] See Box and Jenkins (1976).

3. Another striking property of the residuals is that they expose *time-varying volatility* (volatility clusters). Such heteroscedastic behavior would seem to suggest the application of multivariate conditional volatility models. A pure unconditional stable fitting of the residuals is not able to explain persistence in volatility. Studying the plots of the residuals raises the assumption that the conditional volatility could be contingent on lagged representations of conditional volatility and lagged representations of the residuals.

It is noteworthy that the presence of clustering volatility causes heavy-tailed unconditional distributions even when the conditional distribution is Gaussian. However, heavy-tailed (conditional) distributions and heteroscedasticity models are not mutually exclusive. This discussion is again picked up in chapter 9. However, first an elaborate analysis of observation 1 is given:

As there is a high degree of correlation among the residuals, it raises the assumption that their moves systematically depend on each other. A "common force" seems to drive the residuals. The impact is especially strong for the investment grade ratings, becoming weaker with decreasing credit quality. The residuals can be interpreted as the daily change in credit risk or spread (Khindanova, Rachev and Schwartz, 1999). The moves of the credit spread are driven chiefly by a systematic component due to the strong common behavior of the residuals. The residuals largely show identical signs for all grades. It seems that the residuals of the equations for the AA, A, BBB, BB, and B returns *mainly follow the movements of the residuals of grade AAA*, especially for large representations. Thus, a closer look at whether changes in the credit spread of AAA systematically drive the credit spread of the other rating grades is needed.

8.4.3 The Systematic Credit Risk Component

It can easily be demonstrated that - just as observed for stock markets[6] - the cross-correlations between the residuals of the VECM increase during highly volatile periods. The risk of a given asset portfolio is seen as a result of both volatility and correlation fluctuations. The idea is to choose a *stable one-factor model* that has the ability to capture the essential features of the residuals' cross-correlations between different rating grades in the VECM.

For the credit return model in equation (8.4), the residuals are decomposed into two components:

- A common factor significant for all rating grades.

[6]Longin and Solnik (1999) and Longin and Solnik (1995) argue for the stock market that cross-correlations between stocks actually fluctuate over time, and increase substantially in a period of high market volatility.

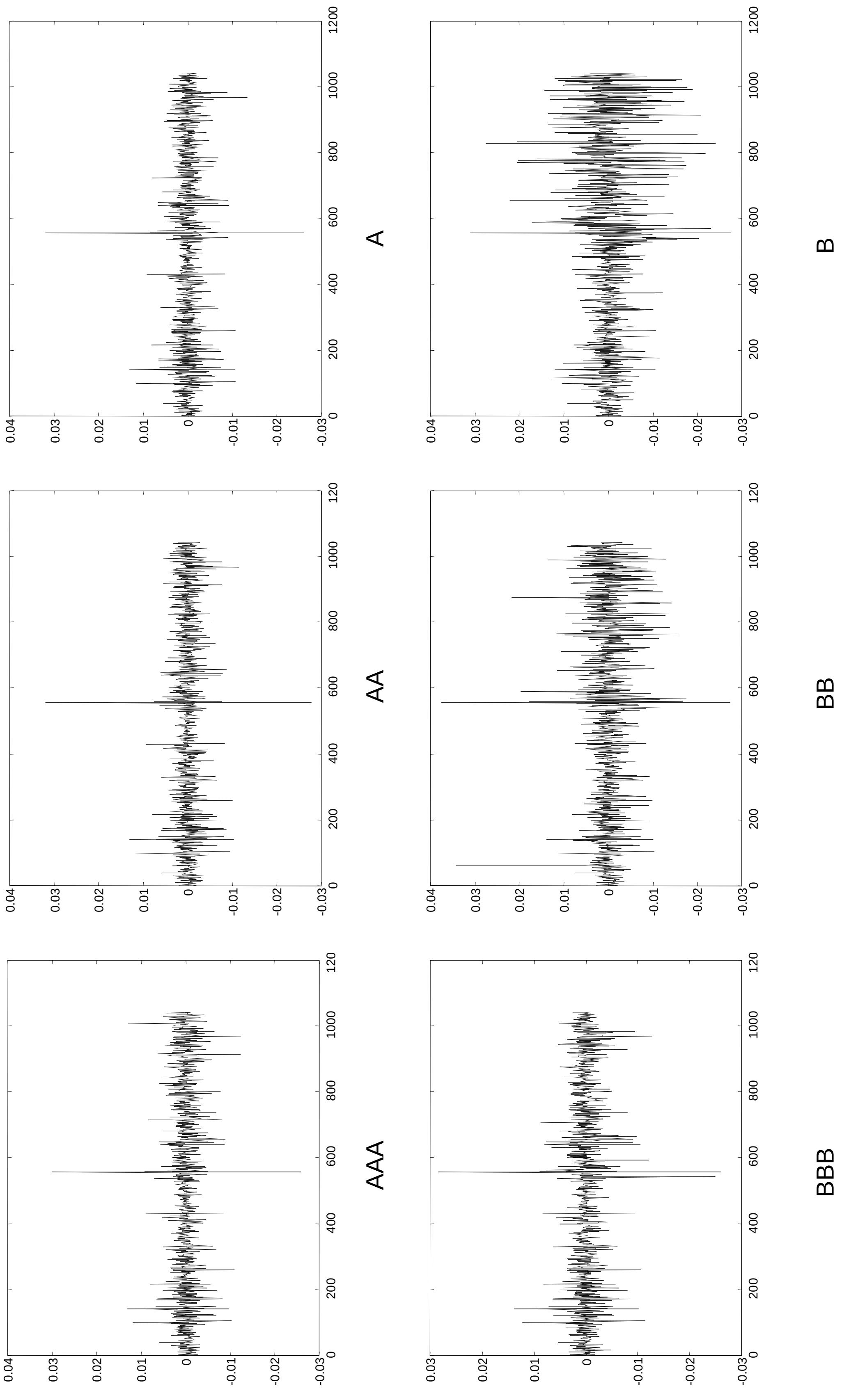

Figure 8.4: Residuals of the VECM for rating grades AAA - B.

- An independent residual part.

In a generic factor model, the residuals are supposed to be combinations of all influencing factors except the chosen common factor and are therefore independent of it. It seems that the residuals of the VECM systematically follow the residuals of the AAA grade equation. Choosing a stable non-Gaussian one-factor model[7], the decomposition of the res_i in (8.4) is performed according to

$$res_{i,t} = d_i res_{c,t} + \epsilon_{i,t} \text{ , where } i = 1, ..., 5. \tag{8.13}$$

$res_{i,t}$ are the residuals of the VECM equations with $i = 1...5$ (for AA, A, BBB, BB, B).[8] $res_{c,t}$ are the residuals of the AAA equation and represent the common factor. ϵ_t is a stable random vector with $1 < \alpha < 2$ with elements $\epsilon_{i,t}$, $i = 1...5$. Both $res_{i,t}$ and $res_{c,t}$ are supposed to have a zero mean, the one-factor model does not have a constant term.

The $\epsilon_{i,t}$ are supposed to be independent of each other. However, if there exist any other common risk factors not captured by the factor model, the $\epsilon_{i,t}$ will not be independent.

The coefficient d_i for the one-factor model is given by

$$d_i = \frac{\sum_{t=1}^{n} res_{i,t} res_{c,t}}{\sum_{t=1}^{n} res_{c,t}^2}. \tag{8.14}$$

In the non-Gaussian case, the coefficient d_i for the one-factor model can also be obtained by ordinary least squares (OLS) estimation. The properties of OLS estimation for stable random variables are discussed in Rachev and Mittnik (2000, p. 90).

The stable one-factor model is an advantageous alternative towards the modeling of innovations with constant correlations. Analyzing the performance of a stable one-factor model, it can be stated that the apparent increase of correlations in high volatile periods can be satisfactorily explained within a simple one-factor model that accounts for heavy tails, although the model remains an approximation (Cizeau et al., 2000). This simplifies the problem of correlation risk and reduces it to a unique factor. However, it is meaningful to have a factor whose volatility is less persistent. In this respect, the use of the AAA residual as the common factor is advantageous. The other residuals of the bond return model show greater persistence in volatility (see also table 8.4).

Thus, within the VAR models, a significant part of the residuals of the equations AA, A, BBB, BB, and B is described by the residuals of the AAA equation. This is one of the key findings in this chapter.

---

[7] Similar to that proposed by Cizeau, Potters and Bouchaud (2000) for stock market returns.

[8] The index i now starts with the AA equation.

8.4.4 Results for the Systematic Credit Risk Component

For the residuals of the cointegrated model, the coefficients of the common factor are obtained by regressing the residuals of the AAA equation over the residuals of the other equations, as stated in (8.13). Aside from other effects, the common factor res_c can be interpreted principally as common change of credit spread over all rating grades.

	d_i	R^2	$t - statistics$
AA	0.7221	0.505	32.59
A	0.5294	0.2915	20.7
BBB	0.5109	0.24	18.13
BB	0.4297	0.054	7.71
B	0.4599	0.0384	6.45

Table 8.1: Coefficients of the common credit risk factor for the 2-year maturities.

	d_i	R^2	$t - statistics$
AA	0.8409	0.7252	41.83
A	0.7904	0.627	52.41
BBB	0.791	0.5862	38.4
BB	0.7888	0.2499	18.62
B	0.778	0.156	13.87

Table 8.2: Coefficients of the common credit risk factor for the 10-year maturities.

	d_i	R^2	$t - statistics$
AA	0.6517	0.4882	31.51
A	0.558	0.3386	23.08
BBB	0.5225	0.2517	18.71
BB	0.3529	0.0413	6.7
B	0.456	0.0451	7.01

Table 8.3: Coefficients of the common credit risk factor for the 30-year maturities.

For the 10-year maturity, the d_i values turn out to be highly significant and vary between 0.78 and 0.84. As tables 8.1 - 8.3 show, the explanatory power (R^2) is very high for the investment grade bonds AA, A, BBB. However, R^2 plummets for the speculative grades. For the 2-year and 30-year maturities, the common factors exhibit less impact. Nevertheless, for the investment grades, R^2 is mostly beyond 0.25. From the results presented in the tables, it can be concluded that the introduction of the common factor is justified. By analyzing the residuals $\epsilon_{i,t}, i = 1...5$, however, it becomes clear that the correlations among the rating grades did not completely vanish, although they have dropped sharply. This might indicate the existence of other common influencing factors. It could also be partially due to the clustering volatility in the $res_{i,t}$ which increases with decreasing credit quality.

The remaining (unsystematic) parts $\epsilon_{i,t}$ of the residuals still exhibit correlation but to a much smaller degree. The systematic component reduces the scale parameter of the remaining residuals $\epsilon_{i,t}$ of the equations for AA, A, BBB, BB, and B. The $\epsilon_{i,t}$ keep the clustering volatility.

As a benefit of the model, the apparent increase of correlations in highly volatile periods can be satisfactorily explained within a simple one-factor model that accounts for heavy tails (Cizeau et al., 2000).

Now both the persistence of the common factor res_c and the persistence of the other res_i are compared. For the common factor $res_{c,t}$, it is preferable that its conditional variance does not exhibit any dependence with the lagged shocks. The process should instead be close to i.i.d. And in fact, when observing the residuals' plots, it seems that volatility clustering increases with decreasing credit quality. Fitting each of the residuals res_i, $i = 1...6$ with a GARCH(1,1) specification, the results for the 10 year maturities are presented in table 8.4.

	ARCH ($res^2_{i,t}$)	GARCH ($res^2_{i,t}$)
AAA	0.422	0.033
AA	0.119	0.728
A	0.194	0.729
BBB	0.129	0.861
BB	0.077	0.842
B	0.044	0.956

Table 8.4: Results of GARCH specification for the $res_{i,t}$.

Table 8.4 clearly evidences that the AAA residuals exhibit the least persistence in variance, having the lowest sum of ARCH and GARCH component. Thus, the clustering volatility is mainly in the residuals of the equations for AA - B. Therefore, the clustering volatility is expected to remain in the $\epsilon_{i,t}$ after subtracting the impact of $res_{c,t}$, given by $d_i res_{c,t}$, from the $res_{i,t}$.

This section has successfully explained the strong dependence and common behavior of the VARs' residuals (observation 1). Observation 2 (heavy-tailedness) and observation 3 (clustering volatility) will be examined in chapter 9. The final section of this chapter analyzes the other common risk factor, the treasury returns.

8.5 The Behavior Of The Treasury Returns

In this section the behavior of the treasury bond returns and log prices is analyzed. There have been examinations in the past on treasury yields and their relation with corporate bond yield spreads, such as (Duffee, 1998). However, in the given case the focus is on the stochastic process of daily log-price changes of treasury bonds (assuming constant time to maturity).

Instead of incorporating the treasury returns as an additional equation into the cointegrated VAR model, the treasury returns enter the model as an exogeneous variable and represent the risk factor for interest risk.

The basic question that arises is "Are the treasury returns x_t in equation (8.4) just a noise or do they depend on the former representations of their log prices or returns?" The objective is to find an appropriate stochastic process that describes x_t.

Intuitively, it would not be appealing to assume x_t as a noise with i.i.d. representations.

The reasons for this are two-fold:

- Historical observations imply for the prices of treasury bonds that - for a given maturity - the interest rates, and thus the prices are most likely to move within a certain range.
- It is reasonable to assume that interest rate and price of treasury bonds move around a long-term mean. The more distant the current price is from the long-term mean, the more likely it is that the price will move in the direction of the mean.

The literature provides a study by Longstaff and Schwartz (1995), who have analyzed the behavior of credit spreads (defined *here* as the difference between a bond's market yield and the treasury yield of equal maturity).

Longstaff and Schwartz (1995) describe the log of credit spread, here expressed by sp_t, to be a mean-reverting process which is represented by an equation

$$\Delta sp_{t+1} = \gamma_0 + \gamma_1 sp_t + \epsilon_t, \tag{8.15}$$

where γ_1 is negative in all regressions. For high-rated bonds the log of the spread is more volatile than for lower-rated bonds.

A look at the log prices of the treasury bonds over time shows that they demonstrate signs of a mean-reverting process too (see figure 8.5). The unit root test reveals that the log prices of the treasuries are I(1) for all maturities.

The relevant regression equation for the mean-reverting process is

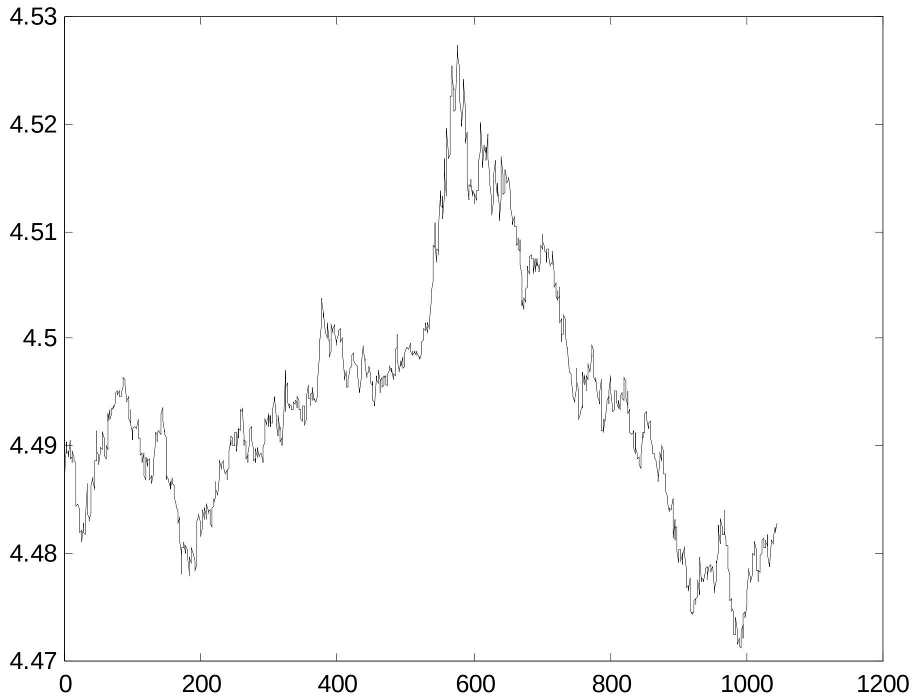

Figure 8.5: Daily log price of 10-year treasury bond over the four-years observation period.

$$x_{t,T} = \Delta y_{t,T}^{(x)} = \mu + \beta y_{t,T}^{(x)} + w_t, \tag{8.16}$$

where $y_{t,T}^{(x)}$ is the daily log price of a zero treasury bond with constant time to maturity T, valued at t. It becomes a mean-reverting process if β is negative. The results of estimating μ and β are presented in table 8.5.

Maturity		Coefficient	Std.-dev.
2 year	μ	0.018175	0.012962
	β	-0.004045	0.002884
10 year	μ	0.022178	0.012559
	β	-0.005502	0.003121
30 year	μ	0.015656	0.00855
	β	-0.005512	0.003054

Table 8.5: Parameters of a mean-reversion process describing daily treasury returns.

The results are significant for all three maturities. The negative sign for the β and the positive sign for the μ clearly indicate the mean-reverting character of the process. Thus, the description of the treasury returns by assuming a mean-reverting process for the treasury log prices is justified. This finding gains significance when simulating future credit returns for the different corporate credit rating grades with the model (8.4). x_t is the risk factor with by far the strongest impact in each equation of the cointegrated VAR. Thus, this influences the price paths of the corporate rating grades also to become a mean-reverting process.

8.6 Conclusion

Chapter 8 has described the specification and fitting of the credit return model for the three maturities 2 years, 10 years and 30 years. It has explained the available data and described the selection of the appropriate lag order, the unit root tests, and the results of the cointegration tests. Moreover, the residuals of the cointegrated VAR models and the treasury returns have been analyzed.

Specifying the cointegrated VAR, one obtains a lag order of 1 for the VECM. It could be observed that the impact of the treasury returns on the corporate bond returns drops with decreasing credit quality. While the treasury bond returns explain roughly 70% of the returns for AAA - BBB grades, their impact on the BB and B falls off strongly.

The chapter comes up with four important results that also have implications on the further course of this thesis:

1. The traditional cointegrating relations obtained by Johansen turned out to be too weak. When simulating future price paths with the cointegrated VAR model, it was found that the Johansen cointegrating relations were not successful in preventing price paths of neighboring rating grades from intersecting. Such intersections, however, are almost never observed in practice. Therefore, a new set of more restrictive cointegrating relations have been developed. These have a rank of 5 (compared to the cointegration rank of 2 obtained by the Johansen test).

2. The residuals of each equation of a cointegrated VAR follow the behavior of the AAA residuals of the same system. The AAA residuals can be interpreted as a *common risk factor* driving the changes in credit spread for each rating grade. The impact of the AAA residuals on the residuals of the other investment grade equations is marked but becomes weaker for the non-investment grades BB and B.

3. The treasury returns are the *strongest common risk factor* within the cointegrated VAR model and follow a mean-reverting process. They are therefore not i.i.d. This knowledge is important when simulating future price paths with the cointegrated VAR models.

4. The remaining residuals $\epsilon_{i,t}$ in the AA, A, BBB, BB, and B equations exhibit both heavy-tailedness and clear signs of heteroscedasticity. The following chapter 9 will further analyze the volatility behavior of the residuals and develop a multivariate model to describe their dynamic volatility under the stable non-Gaussian assumption.

Chapter 9

Dynamic Volatility

This chapter continues the research to describe the behavior of the residuals $\epsilon_{i,t}$.

For this purpose, an appropriate model to capture the heteroscedastic behavior of the credit returns of grades AA - B is developed. As stated in section 8.4.4, the heteroscedasticity is evident in the innovations of the cointegrated VAR credit return model.

First, two alternatives of multivariate models with time-varying volatility are introduced, both under the stable and under the Gaussian assumption. After fitting the models, their empirical performance is compared. A special focus centers on their capability to make accurate Value at Risk forecasts.

Heteroscedasticity has been studied in the stock market for many years. The works of Mandelbrot (1963), Fama (1965a) and Fama (1965b) show evidence that time series of stock market prices are characterized by leptokurtis, heavy-tailedness, and time-variation in conditional variance. As the plots in the previous chapter illustrate, there seems to be a dynamic relationship between the daily representations and volatility of the residuals. Volatility is not constant over time. Bollerslev (1986) and Engle (1982) have introduced models describing such linkages.[1] The models that capture such behavior are generalized autoregressive conditional heteroscedasticity (GARCH) models. For the GARCH(p,q) model, the volatility is dependent on the p prior conditional volatilities and on q previous representations of the time series's squared values. Such autoregressive models for volatility have been extended to the multivariate case as well - for example when describing the returns of different stock markets.

Mandelbrot and Fama found that the volatilities of different securities have common persistent components, which indicates linkages between markets. It is worth examining whether such linkages causing *volatility spillovers* also exist between bond markets with different credit ratings.

[1]See Bollerslev (1992) for an overview.

For the given cointegrated VAR models, the objective is now to isolate the heteroscedastic behavior of the credit returns as part of the residuals, which represent changes in credit spread. Another goal is that the model should be practical and tractable, and not become too complex.

Relevant literature[2] has analyzed the overall returns of an asset for heteroscedasticity. However, thus far research has not been done to attribute this behavior to a certain risk factor. In the given case, the heteroscedasticity is isolated in the disturbance terms of the model. Thus, aside from the traditional approach, the heteroscedasticity is now ascribed to the component that represents the credit rating's individual changes in credit spread.

The chapter is structured in three parts:

- Two multivariate stable volatility models are introduced: the multivariate stable GARCH(1,1) and the multivariate stable exponentially weighted moving average (EWMA) model. Their Gaussian counterparts are also applied for comparison.
- The volatility models are fitted for the residuals $\epsilon_{i,t}$ of the credit return model. Then, the models are compared in terms of their forecast accuracy for volatility.
- For application of the credit return model in a risk-management context, it is essential to check the forecast performance of the multivariate volatility models for VaR.

In this chapter, the cointegrated VAR credit return model is extended to a system that (i) controls the long-term relationships between the different rating grades and (ii) accounts for the short-term movements influenced by the residuals' cross correlations and their common heteroscedastic behavior.

As the main result of this chapter, the *stable multivariate EWMA model* exhibits the best performance for Value at Risk and is therefore chosen for the residuals of the cointegrated VAR model.

9.1 Dynamic Interdependence In A Multivariate Framework

Cross-shocks occur if movements in one equation are affected by past movements in other equations. This causes volatility spillovers. Such interdependence behavior

[2] See, for example, the numerous publications of T. Bollerslev.

has been studied for international stock markets.[3] The examination reveals which markets are influenced by other markets. However, it becomes clear that such lagged interference between different rating grades of the cointegrated VAR is not observable. Rather, simultaneous behavior of the volatilities over the different credit grades is found. Volatility spillovers caused by lagged representations of the volatilities of other rating grades' residuals are not significant. In addition, the lagged credit returns of other grades within the cointegrated VAR are of hardly any significance.

Here each equation of the cointegrated VAR (rating grade) represents a segment of the same national bond market (US Corporate Bonds, industrial sector). So it is not an examination of shock effects between different international markets, and the effect of lagged spillovers does not seem to be likely. Instead, the previous chapter has detected the AAA residuals as a common risk factor for all rating grades. Nevertheless, significant correlations between the (remaining) residuals[4] of the volatility models are still present.

9.2 The Multivariate GARCH Model With Constant Correlation Matrix

The multivariate Constant-Correlation-GARCH (CC-GARCH) bears, as the name says, the restriction of constant correlations. However, compared to multivariate GARCH models with more flexible specifications, it is rather parsimonious.

For the univariate GARCH(p,q) model, the conditional volatility at time t is dependent on the p former representations of the time series's squared values, and on the q former conditional volatilities.[5] The multivariate GARCH framework is a straightforward extension of the univariate case as it is defined as a variance-covariance matrix following a GARCH process. However, the general multivariate GARCH(1,1) model's number of parameters increases dramatically with a growing number of variables (equations). The model with two equations has 21 parameters, three equations have 78 parameters, and five equations have 465 parameters. As the general model is not tractable for five-dimensional multivariate GARCH, the application of the general model is out of question. The remedy is to impose restrictions. One option is to keep the correlation matrix constant. The CC-GARCH(1,1) specification (Bollerslev, 1990) is

[3] See Scheicher (2001) and Isakov and Perignon (2000).
[4] After deducting the impact of the AAA residuals.
[5] The definition of univariate GARCH is presented in section 4.4.

$$\begin{aligned}
\epsilon_{i,t} &= \sqrt{h_{i,i,t}}e_{i,t}, \text{ where } e_{i,t} \sim i.i.d.\ N(0,1), \\
h_{i,i,t} &= a_{i,0} + a_{i,1}\epsilon^2_{i,t-1} + b_{i,1}h_{i,i,t-1}, \\
h_{i,j,t} &= \rho_{i,j}\sqrt{h_{i,i,t}h_{j,j,t}}, \\
\text{with } i,j &= 1...5, \qquad (9.1)
\end{aligned}$$

where $h_{i,i,t}$ is the variance of $\epsilon_{i,t}$, and $\rho_{i,j}$ is a constant correlation coefficient. The indices $i,j = 1...5$ denote the credit ratings AA, A, BBB, BB, and B.

When applied to financial models, the GARCH(1,1) is the most popular of the GARCH specifications (Hull, 2000). Testing the GARCH(p,q) univariate case for each of the $\epsilon_{i,t}$ in the cointegrated VAR for various p and q, the highest likelihood values have also been obtained with the GARCH(1,1) specification. Thus, in the given case the choice for the CC-GARCH is the (1,1)- specification.

The definition in (9.1) refers to the Gaussian case. However, as Rachev and Mittnik (2000) maintain, stable distributions and models with clustering volatility are not mutually exclusive. Although a GARCH process exhibits heavy tails even in the Gaussian case, it has been found from research work with financial time series that - after filtering out the GARCH effects - the remaining process is still heavy-tailed. Thus, it makes sense to allow the GARCH innovations to follow a stable law with $\alpha \leq 2$.

The *stable non-Gaussian* case of CC-GARCH(1,1) also works with constant correlations and is handled analogously. Section 9.4 explains the concept for multivariate ARCH-type models in the stable case. It is constructed via so-called stable subordination.[6]

The correlation matrix of the CC-GARCH is derived from the historical data set and has no predictive property. As previously mentioned and set out in Longin and Solnik (1995), correlations between different financial time series were found to be time dependent. Therefore, it might be desirable to introduce a forecast model that also accounts for conditional correlations.

Another representative of restricted multivariate GARCH models is the so called BEKK model. It was developed by Balsa, Engle, Kraft, and Kroner (Engle and Kroner, 1995) and contains terms for the transmission of volatility shocks, thus allowing interactions between the observed variables while not restricting the model's correlations. However, in the case of five variables, a large number of parameters (exactly 65) still remains. In practice, this type of model just might be applicable for the bivariate case with 11 parameters. For five equations, the BEKK

[6] The specification of univariate stable GARCH is given in Rachev and Mittnik (2000, chapt. 6). However, this concept differs from the approach chosen for the multivariate stable case here using stable subordination.

model is also intractable as the parameters of the cointegrated VAR model have to be considered as well. The CC-GARCH of Bollerslev (1990) is less complex. It incorporates the assumption that all dynamics come from the individual variances themselves.

However, the CC-GARCH can be extended by including spillovers of the other variances in the equation for $h_{i,i,t}$. One obtains

$$h_{i,i,t} = c_{i,i} + a_{i,i} h_{i,i,t-1} + \sum_{j=1}^{k} b_{i,j} \epsilon_{j,t-1}^2, \qquad (9.2)$$

where k is the number of equations. The modified approach considers the impact of volatility shocks that come from the lagged volatilities of other equations.

The constant-correlation GARCH model without extension for cross-volatility shocks has $k(k+5)/2$ parameters.

Given a sample of observations of ϵ_t, the parameters of a multivariate GARCH system are estimated via computing the conditional log-likelihood of the system for each time period and maximizing their sum.[7] The optimization procedure is usually performed with the algorithm developed by Berndt, Hall, Hall and Hausman (1974).

It is necessary and sufficient for covariance matrix H_t if all conditional variances are positive and the constant correlation matrix is positive definite. The BEKK also has these properties but with more parameters. In the BEKK, the covariance matrix is a linear function of its own lagged values and of lagged values of the squared unpredictable returns. It remains to be shown that there are significant conditional correlations. The BEKK could provide conditional correlations. However, with an increasing number of equations there is the danger that the estimation of the model could fail.

Longin and Solnik (1995) and Andersen, Bollerslev, Diebold and Labys (1999) show that, for stock markets, the assumption of constant correlations is violated. In fact, they found increasing correlations for increasing volatility. This was the initial motivation for the BEKK. In contrast to CC-GARCH, the BEKK model is capable of handling dynamic correlations. However, due to the very large number of parameters still remaining in the five-dimensional case, it is not suitable for the credit return model.

The correlation matrix of the CC-GARCH derived from the historical data set has no predictive property. If correlations among different financial time series are found to be time dependent, it is useful to introduce a model that can also forecast conditional correlations.

[7] See Bollerslev and Wooldridge (1992).

In the following section, a suitable multivariate model that stands both for conditional volatilities and for conditional correlations is presented. Moreover, for the purpose of this model, the presence of dynamic correlations between the $\epsilon_{i,t}$ of the credit return model is analyzed.

9.3 The Multivariate EWMA Model

In order to overcome the restriction of constant correlations in the CC-GARCH while keeping a relatively parsimonious multivariate framework for conditional volatilities, the *multivariate exponentially weighted moving average (EWMA) model* is introduced here. It is a frequently used estimator for volatility and correlation.

A simple correlation forecast model such as the historical correlation forecast, mentioned by Lopez and Walter (2000) is considered to be impractical as there is no obvious way to select the interval length.

With the EWMA model, all lagged observations are included. However, current observations have a greater weight than past observations for the calculation of conditional variances and covariances. The model is also present in J.P. Morgan's RiskMetrics$^{(TM)}$.

The EWMA model for volatility is originally defined as

$$h_t = (1-\lambda)\sum_{i=0}^{\infty} \lambda^i \epsilon_{t-i-1}^2, \tag{9.3}$$

with λ as the weighting factor and ϵ_t the representation at time t.

The conditional variance can be reformulated as

$$h_t = (1-\lambda)(\epsilon_{t-1}^2 + \lambda\epsilon_{t-2}^2 + \lambda^2\epsilon_{t-3}^2 + ...). \tag{9.4}$$

$$h_t = (1-\lambda)\epsilon_{t-1}^2 + \lambda(1-\lambda)(\epsilon_{t-2}^2 + \lambda\epsilon_{t-3}^2 + ...), \tag{9.5}$$

with the right term becoming λh_{t-1}.

Thus, the univariate EWMA is expressed as

$$h_t = \lambda h_{t-1} + (1-\lambda)\epsilon_{t-1}^2. \tag{9.6}$$

The EWMA process is equivalent to a GARCH(1,1) process when the GARCH intercept is zero and when $a_1 + b_1 = 1$. Therefore, EWMA can be interpreted as a special case of GARCH in which the persistence parameter is set to unity. For the multivariate case of EWMA, the conditional covariances are calculated as follows:

$$h_{i,j,t} = \lambda h_{i,j,t-1} + (1-\lambda)\epsilon_{i,t-1}\epsilon_{j,t-1}. \tag{9.7}$$

The dynamic (conditional) correlations are obtained by

$$\rho_{i,j,t} = \frac{h_{i,j,t}}{\sqrt{h_{i,i,t}h_{j,j,t}}}. \tag{9.8}$$

As mentioned above, the model is able to forecast conditional correlations as it allows correlations to have a dynamic behavior. The model assumes that observations which are closer to the present have a greater impact on future correlations than those that are further in the past. The degree to which past information influences the forecast is determined by the decay factor $\lambda \in (0,1)$. The proposed value in the J.P. Morgan RiskMetrics $^{(\text{TM})}$ system is 0.94, derived from various empirical analyses. The decay factor does away with the problem of interval building. All observations are included to calculate the conditional correlations; however, the more recent an observation, the greater its weight. Furthermore, by exponentially smoothing out the effect of a change, EWMA correlation forecasts do not exhibit the abrupt changes common to historical correlation forecasts once such a change falls out of the observation period.[8] Unlike GARCH, EWMA does not have a notion of long-run volatility at all and is therefore more robust under regime shifts. Another advantage is its simplicity.

At this point, the thesis work examines whether the observed $\epsilon_{i,t}$ exhibit significant dynamic correlations. Hence, the correlations between the 10-year A and BBB equations are observed. Figure 9.1 plots the conditional correlation obtained with the multivariate EWMA model and compares it with the historical correlation calculated with a moving interval of 100 observations. Both plots exhibit that the conditional correlation increases during highly volatile periods. The volatility can be taken from the plot of $\epsilon_{i,t}$ for the A equation. During volatile periods, large jumps in conditional correlation occur. Due to the relatively short interval, the historical correlation reacts more abruptly to changes and its plot looks very jumpy.

The transition to the *stable* multivariate case of the EWMA is done analogously as for the stable CC-GARCH. The stable multivariate case is performed via the concept of stable subordination, which is explained in section 9.4.

In conjunction with the Gaussian distribution, the EWMA does not allow great jumps. However, due to the heavy weights on more current realizations, the EWMA is able to quickly reflect market shocks.[9] In combination with the stable

[8] Lopez and Walter (2000).

[9] Gibson and Boyer (1999).

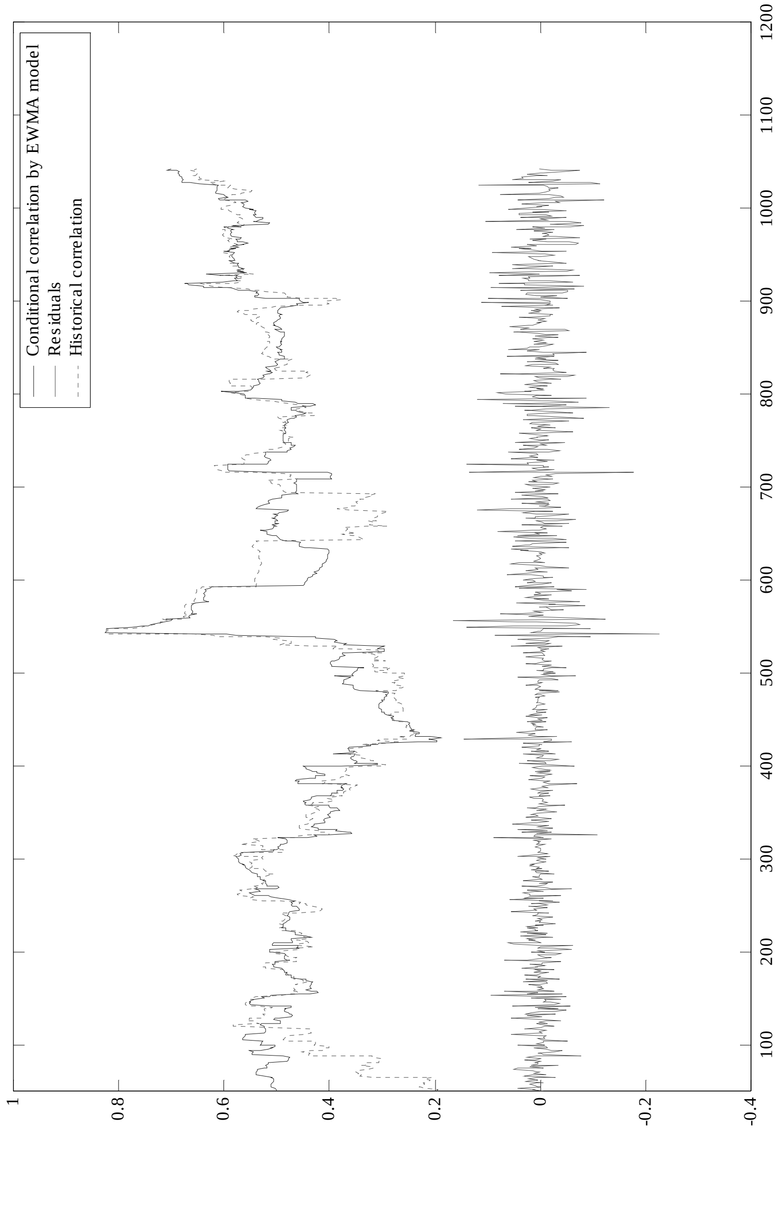

Figure 9.1: Comparison of the EWMA-forecasted conditional correlation with the historical correlation (obtained in a moving interval of length 100) for the 10-year A and BBB equations. In order to illustrate the clustering volatility, the $\epsilon_{i,t}$ for the A equation are plotted as well.

distribution extreme jumps can be modeled more suitably. The application of stable distributions with EWMA and GARCH is more appropriate for considering catastrophe scenarios in the bond market, especially in the junk bond areas, or interventions by central banks. Turmoils in bond markets, as in 1994, are a central issue for risk management and regulatory authorities.[10] The performance of a financial security over a given time is often the result of a few exceptional trading days during the year since other days contribute only marginally to the bottom line.[11] The variance-covariance method based on the normal distribution tends to underestimate the tails of the distribution. A Monte Carlo simulation method based on such an econometric model is subject to model risk. Through application of the stable distribution for the credit model, extreme scenarios can be captured with its tails. Here, the focus is on rare events that cannot be captured by the Gaussian distribution. As the stable method is parametric, it allows out-of-sample stress evaluation for high probability values. Stable GARCH-type models react better to unexpected market shocks and are also somewhat more flexible as they can take skewed conditional distributions into account.[12]

Instead of choosing a proposed value[13] for λ, it is best to fit the proper decay factor for the model by applying maximum-likelihood estimation over the complete multivariate system. The significance can be tested using a likelihood ratio test.

One benefit of the EWMA framework compared to regime-driven models is that volatilities or correlations are not forced to be constant within the regimes. In a regime-driven model, volatilities and correlations can vary solely across regimes. In the regimes approach, the determination of the breaking days poses additional problems.

Analyzed for VaR applications, the Gaussian EWMA was tested to have the lowest capital requirements of various volatility models. Both multivariate GARCH and EWMA performed well in forecasting the covariance matrix.[14] However, EWMA outperforms GARCH in a Value-at-Risk framework.[15]

9.4 Stable Subordination For Multivariate Stable GARCH-Type Models

In order to keep the dependence structure, the stable multivariate GARCH models forecast the variance-covariance matrix, and the concept of stable subordination

[10] Christoffersen, Diebold and Schürmann (1998).
[11] Longin (1999b).
[12] See *skewed stable GARCH* in Rachev and Mittnik (2000, pp. 289).
[13] J.P. Morgan proposes a value of $\lambda = 0.94$ in its RiskMetrics$^{(TM)}$ technical document.
[14] Lopez and Walter (2001).
[15] Covariance forecasts from implied covariance matrices performed poorly in these tests.

is applied.

For dependent random variables following a stable distribution, the variance-covariance matrix does not actually exist. Thus, the truncated variance-covariance matrix is estimated with the multivariate Gaussian volatility model: in this case either with the multivariate EWMA or with the multivariate GARCH(1,1).[16]

$\Sigma_t = \{h_{i,j,t}\}$ is the Gaussian variance-covariance matrix and $h_{i,i,t} = 2\sigma_{i,t}^2$, where $\sigma_{i,t}$ is the scale parameter referring to the Gaussian conditional variance $h_{i,i,t}$.

The concept of stable subordination via the application of a sub-Gaussian vector is demonstrated in the following:

Let $Z \sim S_\alpha(\sigma^*, 0, 0)$ be the vector of a sub-Gaussian $S\alpha S$ random variable. Z can be represented by the product

$$A^{1/2} X = Z, \tag{9.9}$$

where X is a Gaussian random vector with variance-covariance matrix Σ, $X \sim S_2(\sigma, 0, 0) = N(0, 2\sigma^2)$. A is a totally-skewed $\alpha/2$ stable random variable with $A \sim S_{\alpha/2}\left(\frac{\sigma^{*2}}{\sigma^2}\left[cos(\pi\alpha/4)\right]^{2/\alpha}, 1, 0\right)$. X and A are independent.

Two multivariate volatility models based on the stable assumption, the stable CC-GARCH(1,1) and the stable EWMA, have been introduced and their conceptional differences discussed. Furthermore, it has been demonstrated that empirical data shows dynamic correlations.

The following sections introduce measures for analysis of the accuracy of volatility models:

- Statistical loss functions that measure the deviation of the conditional variance forecast from the realized conditional variance are presented (section 9.5.1).
- However, such statistical loss functions have no economic meaning, and as VaR forecasts are the major purpose of the model, it is more important to evaluate their accuracy. Therefore, functions for evaluation of VaR estimates are presented as well (see section 9.5.2).

9.5 Performance Measures For Volatility And Covariance Models

A key criterion for in-sample evaluation of the model fit is the log likelihood. A higher likelihood achieved by the parameters of the system indicates that this

[16] Bravo-Group (2001).

specification is superior to another. The likelihood of a multivariate heteroscedastic model is measured as the sum of the log likelihoods over all equations. For a proper evaluation framework, as described by Lopez, the volatility forecast has to be transformed into a probability forecast. This is not easy for multi-step-ahead volatility forecasts. However, for the given problem, the evaluation is restricted to *one-step-ahead* forecasts.

Originally, volatility forecast evaluation was done by minimizing a *statistical loss function*, for example, mean squared error (MSE). In the past, economic loss functions were unavailable.[17] Most studies have used traditional statistical loss functions to evaluate the performance of volatility models. However, there is a related problem, i.e. squared asset returns were used as a proxy for the latent volatility process. Alternatively, Lopez proposes a framework with a set of loss functions tailored by the appropriate economic interest. However, first the commonly applied statistical loss functions to evaluate the accuracy of variance-covariance models are introduced.

9.5.1 Statistical Loss Function

Mean Squared Error (MSE). MSE is a purely statistical loss function that can be applied to in-sample and out-of-sample testing. It is the average squared difference between the actual conditional variance h_{t+l} and the corresponding forecasted volatility $\hat{h}_{t+l}$. This poses a problem: h_{t+l} cannot be directly observed. Hence, for the calculation of MSE, the squared residuals ϵ_{t+l}^2 are used as a proxy. The multivariate MSE is defined as

$$MSE_{i,j} = 1/n \sum_{l=1}^{n} (\epsilon_{i,t+l}\epsilon_{j,t+l} - \hat{h}_{i,j,t+l})^2 , \tag{9.10}$$

where $\epsilon_{i,t+l}$ and $\epsilon_{j,t+l}$ are the individual residuals and $\hat{h}_{i,j,t+l}$ is the forecast.[18] n is the size of the tested sample.

Mean Absolute Error (MAE). MAE is another typical statistical performance measure that can be applied for volatility models. It penalizes deviation linearly unlike the MSE, which weights large deviation somewhat stronger. The multivariate MAE is defined as

$$MAE_{i,j} = 1/n \sum_{l=1}^{n} |\epsilon_{i,t+l}\epsilon_{j,t+l} - \hat{h}_{i,j,t+l}| . \tag{9.11}$$

[17] Lopez (1999).
[18] Lopez and Walter (2001).

An unattractive property of both MSE and MAE regarding the evaluation of volatility forecasts is that they do not penalize non-positive variance forecasts.

Logarithmic Loss (LL) and Heteroscedasticity-adjusted MSE (HMSE). The logarithmic loss (LL) function with an asymmetric loss penalty function attempts to achieve an improvement compared to MSE and MAE. With the LL function, the penalties are higher when the forecast $\hat{h}_{t+l}$ deviates from lower values of ϵ_{t+l}^2 than from higher values of ϵ_{t+l}^2. However, *the LL cannot be used for covariance forecasts* because these could be negative. Similarly, the heteroscedasticity-adjusted MSE (HMSE) proposed by Bollerslev and Ghysels (1996) allows for asymmetry. The LL is defined as

$$LL_i = 1/n \sum_{l=1}^{n} \left[log(\epsilon_{i,t+l}^2) - log(\hat{h}_{i,i,t+l}) \right]^2 \quad , \tag{9.12}$$

and the HMSE is

$$HMSE_{i,j} = 1/n \sum_{l=1}^{n} \left[\frac{\epsilon_{i,t+l}\epsilon_{j,t+l}}{\hat{h}_{i,j,t+l}} - 1 \right]^2 \quad . \tag{9.13}$$

A property common to MSE, MAE, LL, and HMSE is that they can be applied with no regard for the distributional assumption.

The volatility forecasts considered to be most accurate are those which show the lowest value for a loss function.

However, the major deficiency of the statistical loss functions is their lack of economic meaning. Therefore, it is advantageous to have a performance measure that incorporates economic inference.

9.5.2 Loss Functions With Economic Inference

Forecast evaluation results are greatly contingent on the applied loss function. For a proper evaluation of the volatility forecasts, these have to be transformed into probability forecasts. Therefore, it is important to have a framework that provides the tailored *economic loss function* based on probability forecasts. The forecast of volatility models can be transformed by integration or simulation into probability forecasts of the events of interest. The user then selects a scoring rule and calibration tests over the subsets of the unit interval of interest.

For the given purpose of VaR evaluation, the value of forecasted conditional volatility itself is basically not interesting. According to Lopez, in the case of Value at Risk estimates, the distributional assumptions have a greater impact than the covariance matrix forecasts themselves. Generally, for the accuracy of Value at Risk estimates it is of interest how often the empirical value of the observed asset returns exceeds the forecasted 95% or 99% confidence level of their conditional distribution. Thus, the analysis is based on the counting of exceptions, which are the event of interest. The accuracy of Value at Risk models is determined by the frequency with which the empirical VaR exceeds the estimated VaR.

The task is now to find the volatility model with the best performance in VaR for the credit return model. In the given case, the focus is solely on assessing the performance of the volatility models for ϵ_t as the cointegrated VAR model with all the other risk factors is the same for each of the compared volatility models.

Unconditional coverage. If c is the theoretical confidence level of the VaR (e.g. 95% or 99%), then the unconditional coverage X/n, where X is the number of exceptions and n is the taken sample size, should equal $1-c$.

A conservative view would be minimization of the coverage X/n. However, it is the objective here that the measured unconditional coverage is to reflect the theoretical confidence level of the VaR as precisely as possible.

Along with common practice, one-step-ahead VaR estimates are examined. Applying the definition of VaR[19] to the case of the $\epsilon_{i,t}$ governed by time-varying volatility, one obtains:

$$P(\epsilon_{i,t+1} < -VaR_{i,t}(c)) = P(e_{i,t+1}\sqrt{h_{i,i,t+1}} < -VaR_{i,t}(c)) = 1-c \ , \qquad (9.14)$$

where the expression can also be represented as $P(e_{i,t+1} < \frac{-VaR_{i,t}(c)}{\sqrt{h_{i,i,t+1}}} = 1-c.$[20]

Not only are the number of exceedings but also the magnitude of the exceedings important aspects (Hendricks, 1996; Berkowitz, 1999). However, the latter shall not be considered here.

9.5.3 Evaluation Of VaR Estimates For Unconditional Coverage

Assuming the examined VaR model to be accurate, the exceptions $\epsilon_{i,t+1} < -VaR_{i,t}(c)$ are modeled as independent draws from a binomial distribution. The occurrence of the event $\epsilon_{i,t+1} < -VaR_{i,t}(c)$ should have a probability of $1-c$. Defining $\alpha = 1-c$, this means that there is an unconditional coverage $\hat{\alpha} = X/n$, with n as the sample

[19] The concept of VaR has been introduced in section 2.4.
[20] The VaR is defined as a positive number.

size and X the number of exceptions. In order to evaluate the estimate $\hat{\alpha}$ compared to the assumed significance level α of the VaR, a likelihood ratio statistic is applied:

$$LR_{uc} = 2[log(\hat{\alpha}^X(1-\hat{\alpha})^{n-X}) - log(\alpha^X(1-\alpha)^{n-X})], \quad (9.15)$$

where LR_{uc} has an asymptotic distribution of $\chi^2_{(1)}$. $\hat{\alpha}$ is the unconditional coverage and α is the theoretical significance level of the VaR. This test examines the unconditional coverage of VaR estimates as it counts the exceedings over the entire period. In the bivariate case, the above likelihood ratio statistic is

$$LR_{uc} = 2\left[log(\hat{\alpha}_1^{X_1}\hat{\alpha}_2^{X_2}(1-\hat{\alpha}_1-\hat{\alpha}_2)^{n-X_1-X_2}) - log(\alpha_1^{X_1}\alpha_2^{X_2}(1-\alpha_1-\alpha_2)^{n-X_1-X_2})\right],$$

with

$$[\hat{\alpha_1}, \hat{\alpha_2}] = [X_1/n, X_2/n]; \quad (9.16)$$

The ratio of the binomial case follows a $\chi^2_{(2)}$ structure and the trinomial case has a $\chi^2_{(4)}$.

The evaluation of the credit return model's VaR forecast is performed separately for each equation using the univariate likelihood ratio. A five-dimensional case would be intractable.

VaR estimates are a forecast of the α percent tail of the one-step-ahead return distribution. The evaluation of exceptions can be performed conditionally or unconditionally. If exceptions are counted over the whole period, the test is called unconditional.

Conditional coverage would additionally require serial independence of the exceptions. The testing here is restricted to unconditional coverage. The occurrence of exceptions over the observed sample did not show suspicious clusterings. Thus, it is sufficient to demonstrate the difference in forecast accuracy with unconditional coverage.

9.6 Persistence Of Bond Market Volatility

The volatility of bond market returns and the fact that it is correlated over time has been studied by Fama (1970) and by Bollerslev, Chou and Engle (1992). This was as a rule done empirically; however, not much is known on the theory as yet.

Volatility is interpreted by the information flow.[21] It is assumed that information public to market participants arrives in clusters. Therefore, the volatility is autocorrelated and not independent over time. Jones, Lamont and Lumsdaine (1998) examined the sources of autocorrelated volatility by looking at the response of asset prices to the release of public information. The measurement of how long information is handled in the market is determined by the degree of *persistence*. Jones, Lamont and Lumsdaine have explored the relationship of announcement dates to risk and return for bond markets. Moreover, they have studied day-of-the week effects, showing that the volatility over the week is U-shaped. For example, in their study, Monday and Friday show high volatility whereas Wednesday has the lowest. This is different from French's (1980) study of stock market volatility[22], where he found that the volatility declined over the week.

In a normal univariate GARCH(1,1) model, the persistence in variance is measured by $a_1 + b_1$. For larger values of persistence, but still < 1, a shock in the error term ϵ_t will generate a large variance that remains for a longer time. For values ≥ 1, the process is no longer covariance stationary. It becomes explosive for $a_1 + b_1 > 1$ with a tendency to infinity. For GARCH(1,1), the expectation of the unconditional variance is

$$lim_{s \to \infty} E\left[\epsilon_{t+s}^2 | \Phi_t\right] = \frac{a_0}{1 - (a_1 + b_1)}. \tag{9.17}$$

Φ_t denotes the information at time t which ceases to provide useful information when $s \to \infty$. This formula provides the link between persistence in variance and unconditional volatility. Therefore, increasing $a_1 + b_1$ raises the expected unconditional volatility.

A commonly proposed measure for persistence uses the j-step-ahead forecast of conditional variance,

$$h_{t+j} - \sigma^2 = (a_1 + b_1)^j (h_t - \sigma^2), \tag{9.18}$$

where h_{t+j} is the expectation at $t + j$ of conditional volatility, and σ^2 is the unconditional volatility. The so-called half-line j describes the average time it takes for the conditional variance h_t to revert half-way to the unconditional variance:

$$j = -\frac{log(2)}{log(a_1 + b_1)}. \tag{9.19}$$

In case $a_1 + b_1 > 1$, a negative value for j is obtained.

[21] Roll (1989).
[22] French (1980).

9.7 Forecast Horizon Of Volatility

Christoffersen, Diebold and Schürmann (1998) studied the forecastability of asset return volatilities with regard to the forecast horizon. For short forecasting horizons, traditional dynamic volatility models such as GARCH exhibit strong performance. Andersen and Bollerslev (1997) attest traditional GARCH models as having a good forecasting property for conditional volatility; however, this is different from the so-called end-of-period portfolio value volatility.

In order to obtain volatilities for a long-term forecasting horizon, risk managers used to apply scaling as a popular method. However, its application is not suitable for time-varying volatility models such as the GARCH family since scaling only works in i.i.d. environments. For a GARCH(1,1) environment, the correct conversion from an one-day to an h-day volatility is given by the so-called Drost-Nijman formula, which is very complex. Scaling produces volatilities that are correct on average, but it causes large fluctuations.

In the case of GARCH(1,1), Christoffersen, Diebold and Schürmann (1998) report the volatility as forecastable at all horizons, although forecastability decreases with horizon in accordance with the Drost-Nijman formula (the reader is referred to their paper). However, GARCH modeling may only be an approximation of the true time series process. In order to forecast volatility over longer time horizons, Christoffersen has built a model for the evaluation of interval forecasts. As the VaR measure is known to be the boundary of a one-sided interval forecast, the adequacy of VaR crucially hinges on the volatility dynamics. Christoffersen, Diebold, and Schürmann realize a quick decay of volatility forecastability. Therefore, for long time horizons they would rather focus on direct modeling of extreme tails of return densities. The ability to assess extreme quantiles directly enables extreme risks such as stock market crashes or large bond defaults to be managed. Such great movements are captured by heavy-tailed distributions such as the pareto-stable distribution. The view of the stable GARCH and stable EWMA models held here is the following: both models and their Gaussian counterparts are evaluated exclusively for one-step-ahead forecasts. Theoretically, GARCH models could be scaled in order to obtain a variance-covariance matrix for a longer horizon. This and the use of the stable distribution enhance their long-term forecasting property. The combination of models for volatility dynamics and stable distributions accounts for both volatility forecasting and the capturing of extremal events. However, the objective of this chapter is to demonstrate the dominance of stable multivariate GARCH-type models especially in the one-step-ahead forecasting of VaR.

The next section presents the results of the forecast accuracy of both the stable GARCH(1,1) and the stable EWMA models. The performance measures presented in 9.5 are applied. In the tables, not only the statistical loss functions MSE, MAE, and LL but also the results of the test for unconditional coverage are set out.

9.8 Results For The Stable GARCH And Stable EWMA - Comparison

So far, for a given maturity, the credit returns of different credit qualities (rating grades) have been represented by cointegrated VAR models. Heteroscedasticity is clearly present in the innovations of the cointegrated VAR, which represent the changes in the credit spread. The residual plots in chapter 8 demonstrate the occurrence of volatility clusters.

This section presents the following results:[23]

- The fitting results of the multivariate GARCH(1,1) and EWMA model introduced in this chapter: for each cointegrated VAR (each represents a given maturity), the volatility models have been applied to the ϵ_t. The maturities are 2, 10, and 30 years.
- The predictive accuracy of the forecasted variance evaluated by in-sample tests for both the CC-GARCH(1,1) and the EWMA: for comparison, the traditional statistical loss measures, MSE, MAE, and LL are applied.[24]
- The assessment of the accuracy of both Gaussian and stable VaR: VaR is tested by unconditional coverage $\hat{\alpha} = X/n$ over the sample for one-step-ahead forecasts. The univariate likelihood ratio test LR_{uc} (9.16) is performed for each estimate $\hat{\alpha}$.

The parameters of the GARCH(1,1) are displayed in tables 9.1, 9.7, and 9.13. For the evaluation of EWMA, two values for λ are chosen:

- $\lambda = 0.94$, which is the standard value proposed by RiskMetrics $^{(\text{TM})}$.
- $\lambda = 0.9840$, which is obtained by maximum likelihood estimation.

The comparison of multivariate EWMA and CC-GARCH(1,1) for both the stable and the Gaussian assumptions evidences the predominant performance of the stable distribution in terms of Value at Risk (VaR) forecasting. The 99% stable VaR clearly outperforms the 99% Gaussian VaR because the 99% Gaussian VaR largely underestimates the empirical 99% VaR.

Although the CC-GARCH yields a better log-likelihood value than the EWMA for fitting the variance-covariance matrix, CC-GARCH with Gaussian marginals performs worse in terms of unconditional coverage. This is the especially true

[23] Parts of these results are also presented in Martin and Rachev (2001).

[24] Only the forecasts for the variance but not for the covariance have been evaluated here.

Equation	α_0	α_1	β	MSE (10^{-11})	MAE (10^{-5})	LL	-Logl
AA	0	0.186	0.585	0.0063	0.0118	9.0955	-3.2094
A	0	0.192	0.581	0.0135	0.0157	6.7397	(10^4)
BBB	0	0.143	0.826	0.0534	0.0223	9.0049	
BB	0	0.150	0.910	0.6015	0.0620	10.1862	
B	0	0.034	0.967	0.8422	0.1475	9.6246	

Table 9.1: Stable CC-GARCH(1,1) for 2 Year Maturity: Parameters, Log-Likelihood, and MSE.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.0691	0.0125	0.1401	0.1017
A	0.0605	0.0010	0.1507	0.0979
BBB	0.0566	0.0077	0.0758	0.0393
BB	0.0537	0.0038	0.0355	0.0144
B	0.0701	0.0038	0.0710	0.0288

Table 9.2: Stable GARCH(1,1) for 2 Year Maturity: Comparing Stable and Gaussian Unconditional Coverage for the 95% and 99% Value at Risk.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.9927	0.5650	1.0000	1.0000
A	0.8684	0.9998	1.0000	1.0000
BBB	0.6620	0.5632	0.9996	1.0000
BB	0.4120	0.9787	0.9762	0.8195
B	0.9951	0.9787	0.9967	1.0000

Table 9.3: Stable GARCH(1,1) for 2 Year Maturity: Likelihood Ratio Test for Stable and Gaussian Unconditional Coverage.

Equation	MSE (10^{-11})	MAE (10^{-4})	LL	MSE (10^{-11})	MAE (10^{-4})	LL	-LogL (10^4)
	($\lambda = 0.9840$)			($\lambda = 0.94$)			
AA	0.0059	0.0119	10.7386	0.0059	0.0119	8.3596	($\lambda = 0.9840$)
A	0.0130	0.0155	7.8675	0.0129	0.0157	8.3733	-3.2543
BBB	0.0513	0.0202	10.6148	0.0518	0.0204	10.4092	($\lambda = 0.94$)
BB	0.5189	0.0752	10.1221	0.5247	0.0750	10.6083	-3.2212
B	0.8376	0.1228	10.4895	0.8511	0.1245	11.7623	

Table 9.4: Stable EWMA for the 2 Year Maturity: Log-Likelihood, MSE, MAE, and LL.

for the 2 year and 30 year maturities. The Gaussian EWMA shows much better performance here.

EWMA exhibits better forecasting results for the conditional variance evaluated by MSE and MAE as statistical loss functions. Under LL, CC-GARCH is

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.0585	0.0019	0.0557	0.0221
A	0.0547	0.0019	0.0528	0.0173
BBB	0.0576	0.0058	0.0489	0.0211
BB	0.0633	0.0058	0.0566	0.0259
B	0.0710	0.0058	0.0566	0.0259

Table 9.5: Stable EWMA for 2 Year Maturity: Comparing Stable and Gaussian Unconditional Coverage for the 95% and 99% Value at Risk.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.7803	0.9987	0.5933	0.9993
A	0.5074	0.9987	0.3190	0.9681
BBB	0.7288	0.8608	0.1299	0.9983
BB	0.9419	0.8608	0.6620	1.0000
B	0.9967	0.8608	0.6620	1.0000

Table 9.6: Stable EWMA for 2 Year Maturity: Likelihood Ratio Test for Stable and Gaussian Unconditional Coverage.

Equation	a_0	a_1	b_1	MSE (10^{-8})	MAE (10^{-4})	LL	-LogL (10^4)
AA	0	0.0889	0.9114	0.0038	0.0277	8.8331	-2.4797
A	0	0.0943	0.9291	0.0058	0.0357	8.8002	
BBB	0	0.0811	0.9355	0.0620	0.0506	11.0577	
BB	0	0.2284	0.8708	0.2688	0.2059	11.2862	
B	0	0.0390	0.9702	0.4005	0.3673	12.3725	

Table 9.7: Stable CC-GARCH(1,1) for 10 Year Maturity: Parameters, Log-Likelihood, and MSE.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.0451	0.0038	0.0595	0.0202
A	0.0451	0.0019	0.0393	0.0144
BBB	0.0643	0.0038	0.0480	0.0230
BB	0.0528	0.0038	0.0384	0.0115
B	0.0758	0.0058	0.0537	0.0221

Table 9.8: Stable CC-GARCH(1,1) for 10 Year Maturity: Comparing Stable and Gaussian Unconditional Coverage for the 95% and 99% Value at Risk.

slightly better. However, it remains questionable if - for the purpose of this model - deviations from lower variances should be more heavily penalized than deviations

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.5396	0.9788	0.8291	0.9964
A	0.5396	0.9987	0.9001	0.8199
BBB	0.9581	0.9788	0.2346	0.9997
BB	0.3193	0.9788	0.9265	0.3657
B	0.9996	0.8612	0.4124	0.9993

Table 9.9: Stable CC-GARCH(1,1) for 10 Year Maturity: Likelihood Ratio Test (p-values) for Unconditional Coverage.

Equation	MSE (10^{-8})	MAE (10^{-4})	LL	MSE (10^{-8})	MAE (10^{-4})	LL	-LogL (10^4)
	($\lambda = 0.9840$)			($\lambda = 0.94$)			
AA	0.0038	0.0236	8.8331	0.0038	0.0240	10.5047	($\lambda = 0.9840$)
A	0.0037	0.0333	8.8002	0.0056	0.0334	7.5681	-2.4576
BBB	0.0594	0.0440	11.0577	0.0606	0.0447	10.1062	($\lambda = 0.94$)
BB	0.2127	0.1759	11.2862	0.2148	0.1768	9.6193	-2.4282
B	0.3907	0.3026	12.3725	0.3934	0.3049	10.028	

Table 9.10: Stable EWMA for 10 Year Maturity: MSE, MAE, LL, and Log-Likelihood.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.0499	0.0038	0.0528	0.0202
A	0.0480	0.0029	0.0489	0.0202
BBB	0.0605	0.0048	0.0537	0.0240
BB	0.0614	0.0010	0.0566	0.0278
B	0.0825	0.0048	0.0681	0.0336

Table 9.11: Stable EWMA for 10 Year Maturity: Comparing Stable and Gaussian Unconditional Coverage for the 95% and 99% Value at Risk.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.0118	0.9788	0.3193	0.9964
A	0.2346	0.9934	0.1300	0.9964
BBB	0.8687	0.9397	0.4124	0.9999
BB	0.8978	0.9998	0.6625	1.0000
B	1.0000	0.9397	0.9892	1.0000

Table 9.12: Stable EWMA for 10 Year Maturity: Likelihood Ratio Test (p-values) for Unconditional Coverage.

from larger variances - as caused by LL.

When assessing the performance for VaR, the Gaussian models do not qualify. If a decision in favor of one of the remaining alternatives needs to be made, the stable EWMA is preferred. Not only is its performance for VaR better, its

Equation	α_0	α_1	β	MSE (10^{-6})	MAE (10^{-3})	LL	-LogL (10^4)
AA	0	0.15	0.58	0.0016	0.0259	8.4269	-1.9410
A	0	0.04	0.92	0.0051	0.0376	8.3486	
BBB	0	0.11	0.85	0.0132	0.0511	8.8360	
BB	0	0.10	0.89	0.1397	0.1845	9.3149	
B	0	0.02	0.98	0.2319	0.2753	10.4608	

Table 9.13: Stable CC-GARCH(1,1) for 30 Year Maturity: Parameters, MSE, MAE, LL, and Log-Likelihood.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.0384	0.0029	0.1689	0.1104
A	0.0413	0.0019	0.1046	0.0480
BBB	0.0441	0	0.0720	0.0413
BB	0.0518	0.0019	0.0710	0.0336
B	0.0547	0.0029	0.0672	0.0384

Table 9.14: Stable CC-GARCH(1,1) for 30 Year Maturity: Comparing Stable and Gaussian Unconditional Coverage for the 95% and 99% Value at Risk.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	1.0000	0.5632	1.0000	1.0000
A	0.6970	0.9934	0.9951	1.0000
BBB	0.0118	0.9934	0.2344	1.0000
BB	0.8684	0.9987	0.7288	0.9993
B	0.7803	0.7452	0.5074	1.0000

Table 9.15: Stable CC-GARCH(1,1) for 30 Year Maturity: Likelihood Ratio Test for Stable and Gaussian Unconditional Coverage.

Equation	MSE (10^{-6})	MAE (10^{-3}	LL	MSE (10^{-6})	MAE (10^{-3}	LL	-LogL (10^4)
	($\lambda = 0.9840$)			($\lambda = 0.94$)			
AA	0.0014	0.0215	8.4354	0.0014	0.0217	8.1679	($\lambda = 0.9840$)
A	0.0050	0.0309	8.8667	0.0050	0.0312	8.4267	-1.8895
BBB	0.0134	0.0445	9.7536	0.0132	0.0443	9.2582	($\lambda = 0.94$)
BB	0.1373	0.1530	10.7555	0.1383	0.1555	10.2044	-1.8652
B	0.2326	0.2262	11.3956	0.2343	0.2277	10.9303	

Table 9.16: Stable EWMA for 30 Year Maturity: MSE, MAE, LL, and Log-Likelihood.

simplicity is also a deciding factor. Moreover, it allows for dynamic correlations while CC-GARCH works with constant correlations. The hypothesis of constant correlations is contradicted by the plots of figure 9.1. This is, for example, similar to the findings of Longin and Solnik for stock markets: they state that correlations

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.0518	0.0010	0.0576	0.0230
A	0.0518	0.0048	0.0518	0.0211
BBB	0.0518	0.0058	0.0451	0.0221
BB	0.0653	0.0019	0.0557	0.0192
B	0.0701	0.0058	0.0595	0.0336

Table 9.17: Stable EWMA for 30 Year Maturity: Comparing Stable and Gaussian Unconditional Coverage for the 95% and 99% Value at Risk.

Equation	Stable 95%	Stable 99%	Gaussian 95%	Gaussian 99%
AA	0.2091	0.9998	0.7288	0.9997
A	0.2091	0.9395	0.2091	0.9983
BBB	0.2091	0.8608	0.5392	0.9993
BB	0.9699	0.9987	0.5933	0.9919
B	0.9951	0.8608	0.8287	1.0000

Table 9.18: Stable EWMA for 30 Year Maturity: Likelihood Ratio Test for Stable and Gaussian Unconditional Coverage.

are not constant over time and increase during highly volatile periods.[25]

As the cointegrated vector-autoregressive (VAR) model is built for three different maturities (2, 10, and 30 years), it may be used to simulate future credit returns for all maturities. The returns of maturities lying in the intervals

-] 1 month , 2 years [,
-] 2 years , 10 years [, and
-] 10 years , 30 years [

are obtained by cubic-spline interpolation between the boundaries of the interval. Furthermore, the prices of a one-month bond may be used as fix-points as they show almost no fluctuation compared to the other three maturities.

The risk factors of the credit model are set out below:

- The treasury returns x_t following a mean-reverting process.
- The common credit spread movement $res_{c,t}$.
- The heteroscedastic vector of dependent innovations, $\epsilon_{i,t}$.

[25] Longin and Solnik (1999).

The common credit risk factor and the behavior of the treasury returns have already been discussed in sections 8.4.4 and 8.5.

The proposed simulation framework consists of *three* cointegrated VAR models (maturities 2, 10, and 30 years). It is useful to consider dependence not only between the innovations of the same maturity, but also among the innovations of two different cointegrated VARs. The same applies for both the common credit spread movement and the treasury returns of the different VARs.

9.9 Conclusion

This chapter has developed a solution for describing the multivariate heteroscedastic behavior of the credit returns, which has been isolated in the residuals of the cointegrated VAR.

The results of this chapter can be summarized as follows:

- In order to describe the dynamic volatility of the innovations of the credit return model, the first part has introduced the concepts of multivariate GARCH(1,1) and multivariate EWMA (under both stable and Gaussian assumption). Although GARCH(1,1) is tractable in the five-dimensional case, EWMA has only one parameter to be estimated. The stable cases of GARCH and EWMA are performed via the concept of stable subordination.

- Statistical loss measures to evaluate variance-covariance forecasts are discussed. However, the actual measure of interest is not the volatility itself, but Value at Risk. Value at Risk bears economic meaning as it represents the economic capital required to cover potential credit losses. To evaluate the accuracy of VaR forecasts, a test for unconditional coverage is performed.

- Although the CC-GARCH model achieves a slightly greater log likelihood due to the larger number of parameters, it does not outperform the EWMA model in terms of statistical loss functions. The applied test of unconditional coverage shows that, for the 99% VaR, the stable models are clearly better than the Gaussian models since the latter largely underestimate the empirical 99% Value at Risk. More importantly, the stable EWMA outperforms all other models including the stable CC-GARCH (1,1) in terms of Value at Risk forecast accuracy. The forecasting results, the parsimonious parametrization, and the simplicity of estimation lead to the selection of the multivariate stable EWMA for description of the volatility of the residuals of the credit return model.

- In contrast to CC-GARCH, multivariate EWMA considers time-varying correlations. Figure 9.1 has demonstrated that historical correlations change over time and increase during volatile periods. The plotted correlation estimates of the multivariate EWMA model capture such changes quite well.

However, EWMA and GARCH are "traditional models" in that they deal with processes which have an integer order of integration. In the cointegrated VAR model, the log prices are assumed to be an I(1) process and the returns are assumed to be an I(0) process.

The following chapter covers *fractionally integrated* processes. These processes allow non-integer values for the order of integration and leave behind the traditional jack-knife notion that there are only I(0) or I(1) processes. This restriction excludes a range of possible other processes. Due to the greater flexibility gained by allowing fractional processes, the expected yield is a more precise description of the credit returns and improved forecasting performance of VaR.

Chapter 10

Fractional Models For Credit Data

This chapter introduces the concept of fractionally integrated processes and long memory. Long memory is found in the data of the credit model. The chapter develops an appropriate solution for the description of the credit model's $\epsilon_{i,t}$ by a long-memory process.

Thus far the variables and innovations of the credit return model are assumed to have an integer order of integration: either I(1) or I(0). Such a knife-edge distinction is very restrictive. Allowing fractionally integrated processes results in a better description of the credit return generating processes and, as a consequence, better forecasting results. The basics of fractal processes and long memory have already been presented in chapter 5. Moreover, the returns of corporate bond indices have been analyzed for long-range dependence, and signs of long memory have been found. Now the current chapter applies the concept of fractional integration and long-memory processes in researching for an improved modeling of the $\epsilon_{i,t}$ compared to the traditional volatility model EWMA of the previous chapter.

There are two main objectives this chapter targets:

1. Testing the innovations of the credit return model for fractional integration, which includes both the signed innovations and absolute values of the innovations.[1]

2. If fractional integration is found, a powerful and tractable multivariate model describing the process of the innovations has to be built.

Two possible alternatives are examined under 1:

[1]The absolute values can be seen as a proxy for conditional volatility.

- If fractional integration is significantly present in the signed values of $\epsilon_{i,t}$, then the innovations $e_{i,t}$ of the previously applied GARCH(1,1) and EWMA models should also be tested for fractional integration. A credit return model that keeps its "traditional" multivariate volatility model (GARCH(1,1) or EWMA) is imaginable where the innovations of the volatility model themselves exhibit fractional integration. This kind of approach would easily be tractable in multivariate frameworks with more than two dimensions. Aside from the combination with a classical GARCH or EMWA model, the signed $\epsilon_{i,t}$ should be tested alone for long memory as well.
- Alternatively, the volatility or absolute values of the credit model's innovations exhibit fractional integration.

The findings of this chapter are the following:

- The absolute values of the innovations are driven by a fractionally integrated process which has a long memory. In this case, the above notion of traditional multivariate GARCH or EWMA models with fractionally integrated innovations turns out not to be applicable.
- The absolute values of the credit return model's innovations can best be described using a tractable multivariate FARIMA(p,d,q) process with stable distributions. In this case, the fractional integration is modeled in the credit return model's innovations while the conventional cointegration relations based on I(1) and I(0) are kept.

To be demonstrated in the following chapters is that the chosen long-memory model outperforms the accuracy of the traditional volatility model presented in the previous chapter 9.

An introduction to fractionally integrated processes has already been given in chapter 5 of this thesis. Fractionally integrated processes that are *persistent* are said to have a long memory. The definition and key properties of fractionally integrated processes and long memory are briefly reviewed here.

10.1 Fractionally Integrated Time Series

It is known for Gaussian fractional time series exhibiting long memory that their autocorrelations decrease with increasing lag.[2] The sample autocorrelation function (SACF) exhibits a hyperbolic decay.

[2] The definitions given in this section refer to the Gaussian case.

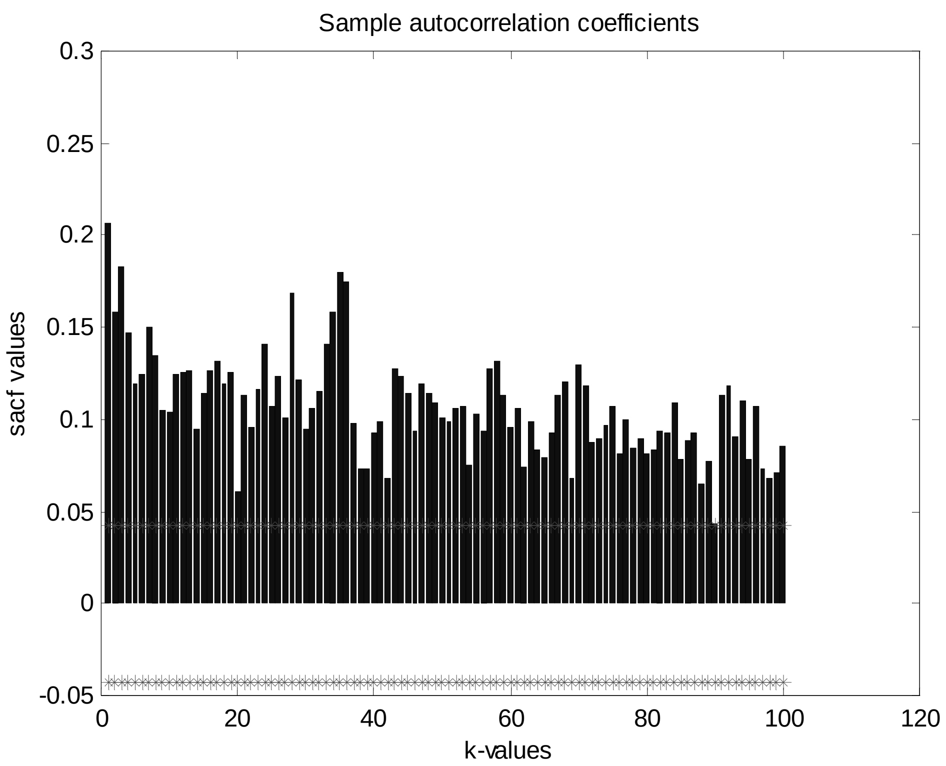

Figure 10.1: Example for an SACF with hyperbolic decay. 'k-values' denotes the lag.

In contrast to a short memory process with geometrically decaying autocorrelation function ACF, a weakly stationary process has a long memory if its ACF $\rho(h)$ has a hyperbolic decay:[3]

$$\rho(h) \sim Ch^{2d-1} \text{ as } h \to \infty, \tag{10.1}$$

where $C \neq 0$, $d < 0.5$.

Baillie (1996) refers to the definition of McLeod and Hipel (1978): a process has a long memory if

$$\lim_{n\to\infty} \sum_{j=-n}^{n} |\rho_j| \tag{10.2}$$

is non-finite. Otherwise it has short-range dependence.[4] A process X_t is said to be integrated of order d, or $I(d)$, if

$$(1-L)^d X_t = u_t, \tag{10.3}$$

where L is the lag operator, $-0.5 < d < 0.5$, and u_t is a stationary and ergodic process with a bounded and positively valued spectrum at all frequencies.[5] If u_t is I(0), i.e. white noise, and covariance stationary, and $0 < d < 0.5$, the process has a long memory.[6] For $-0.5 < d < 0$, the process is *antipersistent* and is said to have an *intermediate memory.*

The above process in (10.3) represents a FARIMA(0,d,0) process. Fractionally-differenced white noise is defined by the representation

$$(1-L)^d (X_t - \mu) = u_t, \tag{10.4}$$

where $E(u_t) = 0$, u_t is i.i.d., and d is possibly non-integer.

The fractional differencing operator $(1-L)^d$, which represents a lag polynomial $1-\alpha(L)$ of infinite order, can be described by

$$(1-L)^d X_t = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} X_{t-j}, \tag{10.5}$$

[3] Brockwell and Davis (1991).
[4] Mikosch and Starica (2000b).
[5] Baillie (1996).
[6] In the Gaussian case.

where $\Gamma(.)$ denotes the Gamma function. Thus, the jth coefficient α_j of the lag polynomial $1 - \alpha(L)$ has the following form:

$$\alpha_j = \frac{-\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} \ ,$$
$$\text{with } \alpha_j \sim \frac{-1}{\Gamma(-d)} j^{-(1+d)} \text{ as } j \to \infty. \tag{10.6}$$

This representation of the α_j coefficients was first seen in Mandelbrot and Ness (1968) and also used by Baillie (1996). Other representations exist, for example Andel (1986), whose polynomials are associated with the Gegenbauer process. Ding and Granger (1996) introduced polynomials defined by the ratio of two beta-functions.

With fractional processes, correlations and weights of the coefficients are characterized by a power law decay. For $d > 0$, partial and inverse correlations also decay hyperbolically.[7]

10.2 Motivation For LRD In Financial Data

As longer time spans of data are available, and more and more high-frequency data is recorded, the basis for alternative volatility estimators and models is created. A more accurate estimation of the inner dependencies of financial time series can be obtained.

Andersen, Bollerslev, Diebold and Labys (1999) address alternative volatility models. For stock markets they found that multivariate aspects of unconditional distribution indicate that the realized covariance tends to be highly skewed, but a simple transformation to correlation delivers normality. Realized correlation is almost always positive. Moreover, realized correlation is often itself highly correlated with realized volatility. Andersen, Bollerslev, Diebold and Labys (1999) call it the "volatility effect in correlation".

In examining conditional volatilities and correlations, Andersen et al. observe a slow hyperbolic decay in the distribution of market volatility. This is a sign for a fractionally integrated long-memory process. A process exhibits long memory if shocks die out at slower than exponential rate, which is characteristic for the persistence of such processes. Baillie, Bollerslev and Mikkelsen (1996) they also obtain a slow hyperbolic rate of decay for the influence of lagged squared innovations of stock market returns using a fractionally integrated GARCH (FIGARCH) model. They argue that the pure distinction between I(0) and I(1) processes is

[7] Peters (1994, p. 192).

too restrictive. Whereas an I(1) process shows infinite persistence, an I(0) process exhibits exponential decay towards shocks. Baillie, Bollerslev and Mikkelsen (1996) state that, for the integrated GARCH of Engle and Bollerslev (1986), the pricing of risky securities may show extreme dependence on the initial conditions. Such extreme dependence could not be observed for the behavior of real prices. With FIGARCH, they have developed a new class of conditional variance which should provide a better explanation for the observed temporal dependencies in the volatilities of financial markets. The influence of the lagged innovations on conditional variance has a slow hyperbolic decay; however, possible shocks to volatility have impulse response weights that tend toward zero.

Long memory builds a bridge between short memory and perfect persistence. Fractionally integrated processes are a framework for time series that show very long dependencies but still seem to be stationary. Another characteristic of long-memory models is that they are supposed to exhibit better forecasting properties for longer horizons than short-memory models with exponential decay of their autocorrelations.

The long-memory property, also known as long-range dependence, describes the high-order correlation structure of a series. In case of long-range dependence, there is a persistent autocorrelation between distant observations. It is a non-linear dependence concerning the first moment of the distribution.

Long memory is present in fractional processes under certain conditions. As set out in chapter 5, there are two different types of fractional processes: *persistent* and *antipersistent* ones. Persistent fractional processes exhibit long-range dependence and have a fractional differencing parameter d with $0 < d \leq 0.5$, whereas antipersistent processes, which are said to have an intermediate memory, have $-0.5 \leq d < 0$.[8]

Evidence as to whether long memory is present can be obtained by direct estimation of the long-memory parameter d. In order to determine the fractional differencing operator d and, thus, the degree of long memory in the data, several procedures have been developed with time, e.g. maximum likelihood estimation procedures, and spectral regression methods (see Geweke and Porter-Hudak, 1983). A short review of the common estimators and a description of the one used in this context follow in a later section.

This chapter examines whether and to what degree the credit return model's innovations ϵ_t, which have been described by the EWMA model in the previous chapter, exhibit long-range dependence. If long memory is present, this could allow a more precise description of the credit return model.

As known from chapter 5, long memory in the returns of indices of corporate

[8]Under the Gaussian assumption.

bonds could also be confirmed under the stable non-Gaussian assumption.[9] The following section now analyzes the $\epsilon_{i,t}$ for the presence of LRD in both signed values and absolute values.

10.3 Testing For LRD In The Data

Recall the credit return model whose innovations $\epsilon_{i,t}$ have been described by a multivariate GARCH(1,1) or EWMA process in chapter 9:

$$\Delta y_{t,T} = \Pi y_{t-1,T-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j,T-j} + B x_{t,T} + res_t,$$
$$res_{i,t} = d_i res_{c,t} + \epsilon_{i,t} \text{ , where } i = 2 \ldots 6, \text{and } res_{c,t} = res_{1,t}, \tag{10.7}$$

where p is the order of the unrestricted VAR, $\Delta y_{t,T}$ is the vector of returns, $y_{t-1,T-1}$ is the vector of lagged log prices, and $x_{t,T}$ are the treasury returns. The matrix Π can be decomposed by $\Pi = \alpha\beta'$. res_t is a six-dimensional, symmetrically stable distributed random vector with elements $res_{i,t}$, $i = 1 \ldots 6$, for the AAA, AA, A, BBB, BB and B grades, and stability index $1 < \alpha < 2$.

$res_{i,t}$, $i = 2...6$, are the residuals of the VECM equations for the AA, A, BBB, BB, B grades. $res_{c,t}$ are the residuals of the AAA equation and represent the common factor.

In chapter 9 the $\epsilon_{i,t}$ have been described applying a five-dimensional multivariate EWMA or GARCH(1,1) model.

In order to acquire information about the existence of long memory in the $\epsilon_{i,t}$ of the credit return model, a number of alternative choices are investigated here:

1. The innovations $\epsilon_{i,t}$ of the model follow a traditional volatility process, GARCH or EWMA, but the noise of the GARCH or EWMA process is not i.i.d. and exhibits long memory.
2. The signed innovations $\epsilon_{i,t}$ of the credit model exhibit long memory.
3. The signed innovations $\epsilon_{i,t}$ do not show long memory, but the absolute values of $\epsilon_{i,t}$, $|\epsilon_{i,t}|$, have a long memory. Thus, the LRD is in the volatility.

The fact that a long memory has been found in the credit returns in chapter 5 raises hope for alternatives 1 and 2 above.

[9] See also Martin, Rachev and Schwartz (2002).

Alternative 1 would be a new kind of model: a combination of a traditional GARCH-type volatility model with a fractional noise. This could be a promising and tractable framework for a multivariate approach.

Martin, Rachev and Schwartz (2002) have demonstrated that credit return series can exhibit significant LRD in the returns.

However, when testing the log returns of the credit model, the $\epsilon_{i,t}$, and the noise of the GARCH/EWMA for long-range dependence, results showed that only the two speculative credit grades BB and B demonstrate signs of LRD (d is found around a value of 0.05).

Instead, the application of the MRS statistic[10] indicates by far a higher significance for long memory in the absolute values of $\epsilon_{i,t}$ (representing conditional volatility) than in the signed $\epsilon_{i,t}$ or in the noise of the traditional GARCH / EWMA model. Figures 10.2 , 10.3, and 10.4 present the MRS statistic for some selected $|\epsilon_{i,t}|$. In all cases, the value of the statistic reaches levels that exceed 19 by far. This indicates long-range dependence at a confidence level of more than 95%. In fact, for all chosen samples, the value lies beyond 30.

This is a remarkable result. Certainly, while there may be significant LRD in the signed $\epsilon_{i,t}$ or GARCH/EWMA noise, d usually has only very low positive values in this case.

Thus, long memory is detected in the volatility represented by the absolute returns.

This result is in line with the findings of Mikosch and Starica (2000b) who also confirm that the log returns of financial time series might only show slight long-range dependence, even when it is significant. LRD is usually measurable more noticeable in the volatility (represented by absolute returns in the stable non-Gaussian case) of the price process. Of course, a pattern such as long memory in the data of absolute returns is also present in the signed returns themselves. Yet, the stochastic process which determines the sign makes the long-memory effect fade for the signed returns.

Thus, it is necessary to focus directly on the process that causes the long-range dependence and to develop a long-memory model for the volatility.

The focal point now targets long-memory models describing volatility. The following section provides a brief literature review of Gaussian long-memory models, describing conditional volatility in the univariate case. Later the models are discussed in the multivariate case.

First univariate long-memory volatility models are discussed (LM-ARCH, FI-GARCH). Next, the multivariate cases of these models are described.

Finally, an explanation is provided as to why these multivariate volatility models are not applicable for the given credit return model. Instead, a multivariate

[10] Introduced in chapter 5.

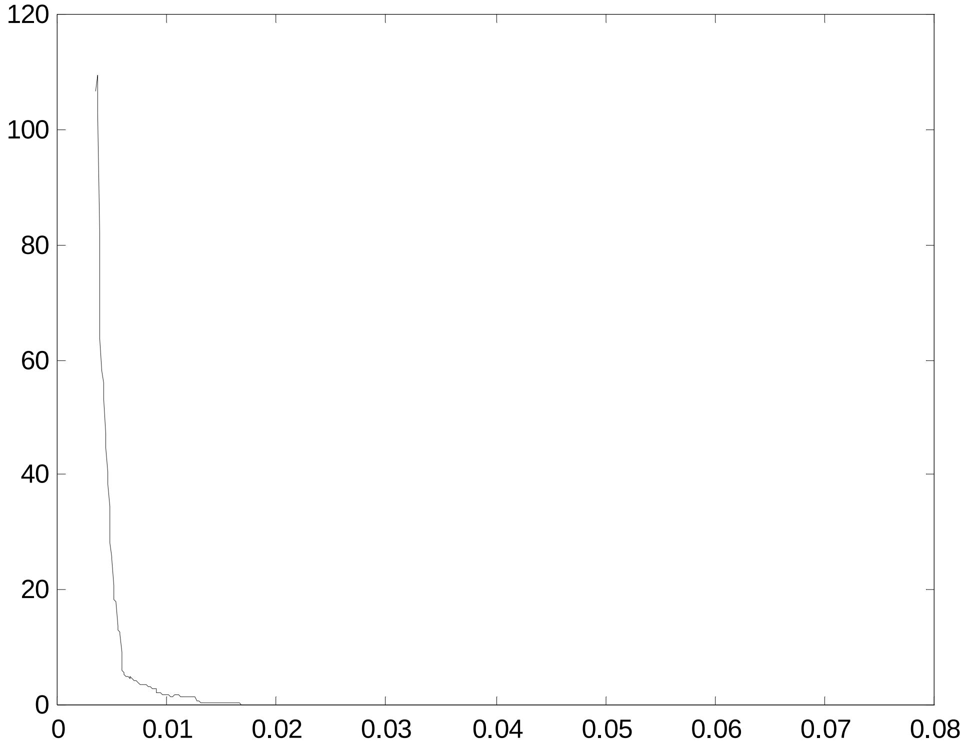

Figure 10.2: Plot of $\theta W(\theta) - \theta$: MRS statistic for absolute values of $\epsilon_{i,t}$ for AA with 30-year maturity.

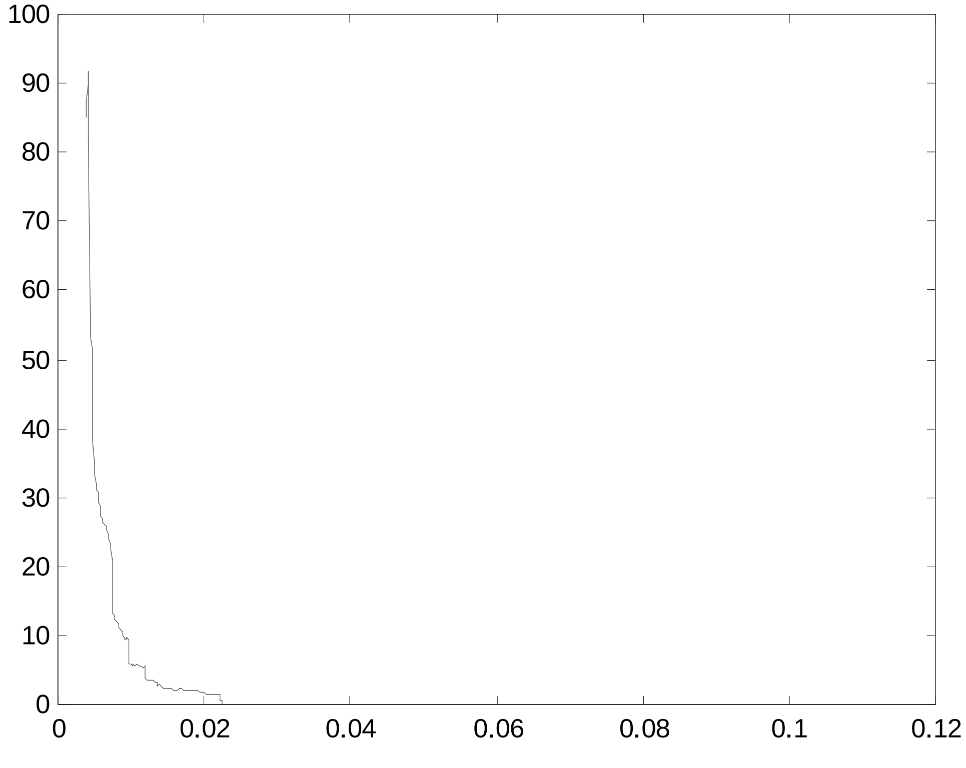

Figure 10.3: Plot of $\theta W(\theta) - \theta$: MRS statistic for absolute values of $\epsilon_{i,t}$ for BBB with 30-year maturity.

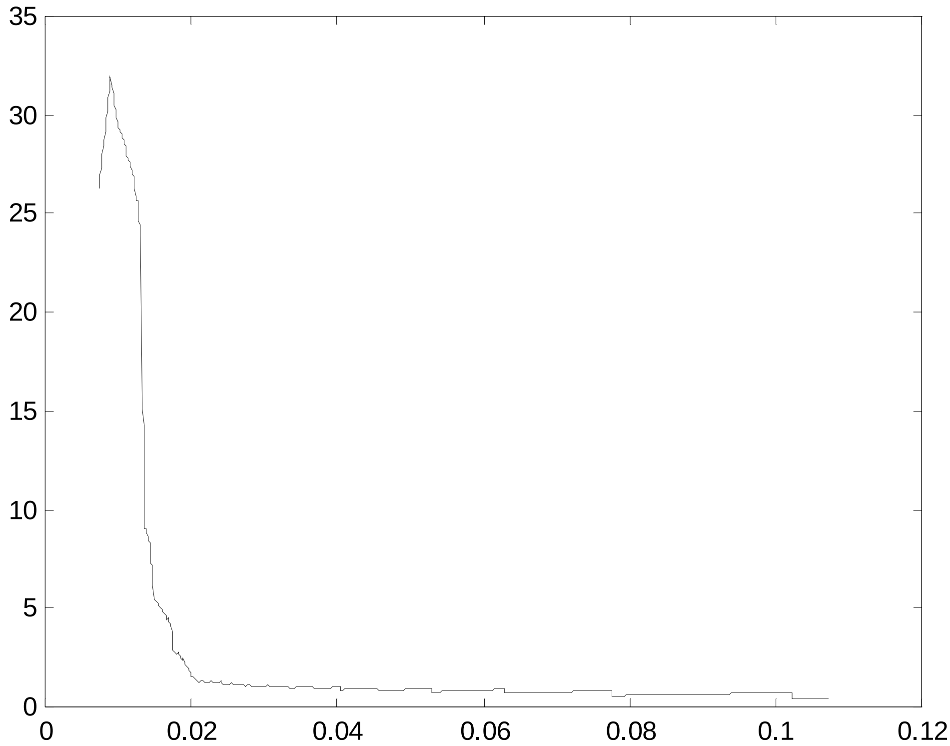

Figure 10.4: Plot of $\theta W(\theta) - \theta$: MRS statistic for absolute values of $\epsilon_{i,t}$ for B with 30-year maturity.

fractional ARIMA process is chosen to forecast the absolute values of the $\epsilon_{i,t}$.

10.4 Models For Long Memory In Volatility

This section covers known long-memory conditional volatility models under the Gaussian assumption. It starts with a basic form, the LM-ARCH, and explains the main features of its general form, the FIGARCH.

According to Mikosch and Starica (2000b), empirical work has shown that a simple ARCH(p) process has a reasonable fit to real financial data only if the number of parameters is rather large. This has been an incentive for the notion of GARCH(p,q) processes. However, traditional GARCH models themselves are not designed to capture LRD in volatility.

GARCH models can fit financial returns only for relatively short periods of time[11] and have to be updated regulary due to changes in parameters. Because of the inability of ARCH(p) and GARCH(p,q) to explain LRD, a more general framework with infinite number of parameters has been introduced. The common form of LM-ARCH has a conditional variance with long-range dependence and is represented by

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j X_{t-j}^2. \tag{10.8}$$

Models of this type have been introduced, among others, by Baillie, Bollerslev and Mikkelsen (1996) and Ding and Granger (1996). It has a long-memory infinite order lag polynomial and its coefficients α_j are strictly positive. This model is a representative of more general models, so-called *fractionally integrated GARCH (FIGARCH)* models. FIGARCH embraces the idea of applying fractional integration to the variance of a process. Although traditional GARCH models account for persistence in volatility, persistence decays relatively fast. In contrast, integrated GARCH (IGARCH)[12] shows a strong impact of volatility shocks. Shocks to conditional variance never die out here. The impact is usually stronger than can actually be observed for real data. FIGARCH is the bridge between GARCH and IGARCH. Thus, the GARCH(1,1) and EWMA[13] models applied in an earlier chapter mark two extremes in terms of persistence.

There are two theories for the presence of persistence in volatility. Lamoreux and Lastrapes (1990) claim that structural breaks are responsible for this phe-

[11] See also Christoffersen, Diebold and Schürmann (1998) and Mikosch and Starica (2000b).
[12] IGARCH is a form of GARCH(1,1) where a_1 and b_1 sum up to 1.
[13] EWMA is a form of IGARCH.

nomenon. Another explanation is that persistence in trading volume causes persistence in volatility. Long memory in trading volume and in the increments of trading time was analyzed for stocks and foreign exchange rates by Marinelli et al. (1999, 2000) and Brunetti and Gilbert (2000).

As for EWMA, the FIGARCH process is strictly stationary even though it is not covariance stationary. Andersen and Bollerslev (1997) consider the long memory of volatility to be an intrinsic feature of the return generating process.

The FIGARCH defined in Baillie, Bollerslev and Mikkelsen (1996) is formulated similar to a FARIMA (*fractionally integrated autoregressive moving average*) process.

Their FIGARCH(p,d,q) is represented by

$$\begin{aligned} \phi(L)(1-L)^d X_t^2 &= \alpha_0 + (1-\beta(L))\nu_t \ , \\ \nu_t &= X_t^2 - \sigma_t^2, \end{aligned} \tag{10.9}$$

where $\phi(L)$, $(1-\beta(L))$ are polynomials in the lag operator L with zeros outside the unit circle, $d \in [0,1]$ and $\alpha_0 > 0$. The FIGARCH model is a generalization that switches to a GARCH model when $d = 0$ and to an integrated GARCH (IGARCH) model when $d = 1$. Baillie, Bollerslev and Mikkelsen (1996) found slowly decaying autocorrelations for the S&P 500. The conditional variance of the FIGARCH is then expressed as

$$\begin{aligned} \sigma_t^2 &= \frac{\alpha_0}{1-\beta(1)} + \lambda(L)X_t^2 \ , \\ \text{where } \lambda(L)X_t^2 &= 1 - \left\{ \left[\Phi(L)(1-L)^d\right] / \left[1-\beta(L)\right] \right\} . \end{aligned} \tag{10.10}$$

Baillie, Bollerslev and Mikkelsen (1996) consider their FIGARCH stationary and ergodic for $0 < d < 1$, the persistence allowed in the variance is larger than in the mean. If all the roots lie outside the unit circle, then the X_t^2 are a stationary process.

Comparing the FIGARCH with a FARIMA process, it has to be noted that FARIMA has no sign restriction whereas for FIGARCH it has to be ensured that all conditional variances are non-negative.

However, Mikosch and Starica (2000b) point out that, for LM-(G)ARCH models, the issue of stationarity has not yet been clarified. According to them, the definition in (10.9) is only valid if the existence of a stationary solution can be proved for the following representation:

$$\sigma^2 = \frac{\alpha_0}{1-\beta(1)} \left(1 - \frac{\Phi(L)(1-L)^d}{1-\beta(L)} \right), \tag{10.11}$$

together with a stationary solution for $X_t = \sigma_t Z_t$.

Therefore, the FIGARCH needs some additional constraints if stationary solutions are to be obtained (see Giraitis, Kokoszka and Leipus, 2000). Clarifying the stationarity issue is important since LRD effects might be due to non-stationarity.

Compared to LM-ARCH, FIGARCH is more flexible, however it takes a greater number of restrictions to keep the conditional variance strictly positive. It follows from the results of Bollerslev and Mikkelsen (1996) and is presented by Brunetti and Gilbert (2000) and Teyssiere (2000). Positive definiteness is ensured for FIGARCH(1,d,0) and FIGARCH(1,d,1) by imposing the following restrictions:

FIGARCH(1,d,0) requires

$$\beta \leq d \leq 1 \text{ and } \alpha_0 > 0. \tag{10.12}$$

FIGARCH(1,d,1) requires more restrictions:

$$\begin{aligned} \alpha_0 &> 0, \\ \beta - d &\leq (2-d)/3, \\ d(\phi - (1-d)/2) &\leq \beta(\phi - \beta + d). \end{aligned} \tag{10.13}$$

Thus, it can be seen from these examples that even the estimation of a simple FIGARCH model becomes very complex, and this already in the univariate case.

The following section explains the multivariate forms of LM-ARCH and FIGARCH.

10.5 Multivariate LRD Models

First, the representation of the multivariate LM-ARCH with constant conditional correlations is introduced:

$$\begin{aligned} s_{i,i,t} = \sigma^2_{i,i,t} &= \sum_{k=1}^{\infty} \alpha_k X^2_{t-k} \text{with } i = 1, ..., n \\ s_{i,j,t} &= \rho_{i,j}\sigma_{i,i,t}\sigma_{j,j,t} \\ \text{with } i, j &= 1, ..., n \text{ and } i \neq j. \end{aligned} \tag{10.14}$$

The conditional covariance matrix is always positive definite. The multivariate long-memory ARCH process is stationary if all the univariate processes on the main diagonal are stationary.[14]

[14]See also Bollerslev (1990).

For the multivariate FIGARCH, a setting with constant correlation matrix is proposed as well.[15] The multivariate constant-correlation FIGARCH is described by

$$s_{i,i,t} = \sigma^2_{i,i,t} = \frac{\alpha_0}{1-\beta_i(1)} + \left(1 - \frac{(\phi_i(L))(1-L)^d}{1-\beta_i(L)}\right) X^2_{i,t} \ , \quad \text{with } i = 1, ..., n.$$
$$s_{i,j,t} = \rho_{i,j}\sigma_{i,i,t}\sigma_{j,j,t} \ , \quad \text{with } i,j = 1,...,n \text{ and } i \neq j. \tag{10.15}$$

The constant correlation parametrization is the most parsimonious among all multivariate FIGARCH specifications (similar to CC-GARCH). Furthermore, under weak conditions the variance-covariance matrix is positive definite. The conditions for positive definiteness are given in (10.12) and in (10.13). Due to the constant correlations, the stationarity is ensured for the multivariate process if the elements on the main diagonal are stationary.

Brunetti and Gilbert (2000) propose a bivariate framework for FIGARCH with fractional cointegration.

It combines the multivariate GARCH with univariate FIGARCH models. The long-term dependence between two closely related markets is captured by a fractional cointegration model.

In a multivariate environment, if the observed markets show long memory in volatility, this model requires different markets to exhibit a common order of fractional integration. Only if there is a common order of integration, it makes sense to ask if the time series are cointegrated.

It could be that the multivariate volatility processes are independent but are driven by a common information arrival process. Therefore the question rises if the processes exhibit a common order of integration (Brunetti and Gilbert, 2000). There could exist a long-run linear relationship between the volatilities which has a lower order of integration than the volatility processes themselves. If the price movements of the different markets have a common source of information then they are supposed to have a common order of integration.

As known from chapter 6, in order to establish a linear cointegrating relationship, a common order of integration is a necessary condition.(Abadir and Taylor, 1999; Robinson and Marinucci, 1998)

The general definition of cointegration (Engle and Granger, 1987; Granger, 1986) also includes the fractional case:

- Two time series, X_t and Y_t, are fractionally integrated of order (d,b) if $Z_t = (Y_t - \beta X_t) \sim I(d-b)$, where $d > 1/2$ and $d \geq b > 0$.

[15] Brunetti and Gilbert (2000).

- The standard test for fractional cointegration requires the determination of the orders of integration for both series under consideration, d_1 and d_2. In order to be able to continue, both series must have a common order of integration. Thus, a hypothesis test $d_1 = d_2$ has to be performed.
- If $d_1 = d_2$ cannot be rejected, the linear cointegrating relationship is estimated and the order of integration d' of the cointegrating residuals has to be determined.
- In case $d = 1$ was found for both cointegrated series, d' gives a test for fractional cointegration. The OLS estimation of the cointegrating vector is only consistent for $d > 1/2$. [16]
- If $d \neq 1$, a further step is required which performs a test for $d' < d$.

The error correction for a fractionally cointegrated model is relatively complex. The reader is referred to Baillie (1996) or Brunetti and Gilbert (2000).

This section has provided an overview on multivariate long-memory volatility models discussed in the literature. However, for a number of reasons they are *not suitable* for modeling the conditional volatility of the given credit return model:

- The presented long-memory conditional volatility models are relatively complex in the multivariate case, especially for the FIGARCH model. (For example, five-dimensional multivariate FIGARCH for the innovations of the returns including a replacement of the model's existing I(0)-I(1) cointegration relation by fractional cointegration is not tractable.)
- The FIGARCH with fractional cointegration increases complexity and it is only applicable if there is a common order of integration.
- The long-memory conditional volatility models are defined for the Gaussian case. An application in a multivariate environment with stable distributions would require a concept such as stable subordination that would increase the complexity and worsen tractability.

Due to the complexity and intractability of high-dimensional multivariate cointegrated FIGARCH models, a new unconventional approach is seeked to handle cointegration, multivariate dependence and long memory of the cointegrated credit return model.

Moreover, fractional cointegration can only be implemented if different credit ratings have identical fractional differencing parameters d. For this reason, it is

[16] See Robinson and Marinucci (1998).

reasonable for to *keep the traditional integer cointegration model* for the long-term relationship, but *allow each equation to have a different fractional differencing parameter* for the LRD model.

As conditional variance does not exist in the stable non-Gaussian case it is preferrable to directly model the absolute values of the $\epsilon_{i,t}$, $|\epsilon_{i,t}|$.

A *multivariate fractionally integrated autoregressive moving- average (FARIMA) model* provides the desired properties. The following section reviews the properties of Gaussian FARIMA. The stable FARIMA follows afterwards.

10.6 The Gaussian FARIMA

The FARIMA (p,d,q), process has a fractionally integrated conditional mean. The standard Gaussian FARIMA process of order (p,d,q) with mean μ may be written in backward-operator notation:

$$\Phi(L)(1-L)^d(X_t-\mu) = \Theta(L)u_t \ , \quad \text{with } u_t \sim \text{i.i.d. } N(0,\sigma_u^2), \tag{10.16}$$

where L is the *backward operator* or *stable operator* and $(1-L)^d = \Delta^d$.[17] The u_t follow a Gaussian law. Integer values of d would lead to an ARIMA process. FARIMA generalizes the class of ARIMA models to *non-integer values* of d. Like an ARIMA(p,d,q) model, the FARIMA(p,d,q) can be differenced, in case $d \geq 0.5$, to obtain a FARIMA(p,d-1,q) process.

A FARIMA(p,d,q) process includes short-memory AR or MA processes over a long-memory process. According to E.E. Peters (1994), those properties provide the potential to describe financial markets.

For the defined Gaussian case, the process is both stationary and invertible if all roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle and $|d| < 0.5$.[18] For $d > 0.5$, the process is non-stationary as it possesses infinite variance. For $d < 1$, the process is mean-reverting. The autocorrelations of the FARIMA process show hyperbolic decay for high lags: $\rho_k \approx ck^{2d-1}$.

For $d \in (0.5, 1.5)$, the differenced time series ΔX_t will be stationary, with intermediate memory for $d < 1$ and long memory for $d > 1$.

According to Baillie (1996), the impulse response function of FARIMA is obtained by first differencing of the above equation (10.16):

$$AL(u_t) = (1-L)X_t = (1-L)^{1-d}\Phi(L)^{-1}\Theta(L)u_t. \tag{10.17}$$

[17] Please recall section 4.3 for non-fractional ARIMA processes.

[18] Invertibility comes from $d > -1$.

The impulse response function describes how endogeneous variables respond over time to a one-period shock in that variable and every other endogeneous variable. The function traces the behavior to shocks (Pindyck and Rubinfeld, 1991, pp. 385). This shock filters through all following periods.

The impulse response function $I(l)$ can be thought as the effect of a shock of size *one* at time t on following representations X_{t+l}. The impulse responses for a stationary process are the coefficients of its *infinite moving average* representation. For FARIMA(0,d,0), the MA representation is given by

$$y_t = (1-L)^{-d} u_t = \sum_{j=0}^{\infty} c_j u_{t-j},$$
$$\text{with } c_j = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}, \tag{10.18}$$

which demonstrates the decay of a shock to such a process.

For FARIMA(0,d,0), the impulse response is represented by:

$$A(L)u_t = (1-L)X_t = (1-L)^{1-d} u_t. \tag{10.19}$$

The prediction from a FARIMA(p,d,q) process is given in Granger and Joyeux (1980) and Geweke and Porter-Hudak (1983) by using the infinite autoregressive representation:[19]

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + u_t,$$
$$\text{where } \pi(L) = (1-L)^d \Phi(L) \Theta(L)^{-1}. \tag{10.20}$$

An important implication from the differences in coefficient decay rates is that fractionally differenced models provide better forecast results compared to the knife-edge choices I(0) and I(1). Between fractionally differenced models and I(0) models, the rate at which past information is useful in forecasting future values differs significantly. For example, comparing an AR(1) with an AR(d) process, the first autocorrelations of both series are almost identical.[20] After that, the autocorrelations of the fractional series fall slowly while the AR(1) autocorrelations fall rapidly which signals the difference between exponential and hyperbolic decay. When comparing AR(1), FARIMA(1,d,0) and FARIMA(0,d,1), it shows that

[19] See also Baillie (1996).

[20] See Peters (1994, pp. 194-195).

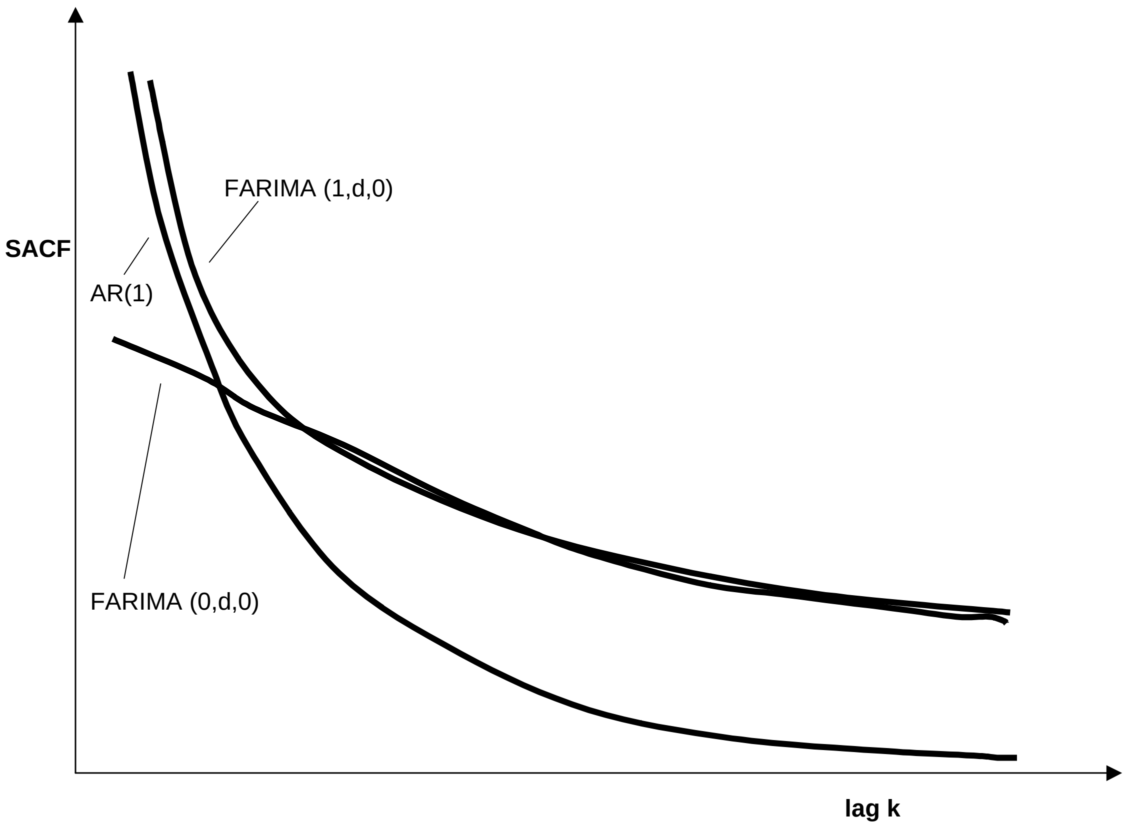

Figure 10.5: Sample autocorrelations for AR(1), FARIMA(0,d,0), and FARIMA(1,d,0).

FARIMA(1,d,0) and AR(1) have almost identical decay for the first autocorrelations, followed by more rapid decay of AR(1). FARIMA(0,d,1) starts with more rapid decay than FARIMA(1,d,0). After that its autocorrelations increase for some lags and move towards FARIMA(1,d,0). For larger lags, the autocorrelations of both processes move in parallel ways. Even high order ARMA processes show a more rapid decay than fractional ARMA processes.

Hosking (1981) suggested a procedure to specify a FARIMA(p,d,q) model for a time series Y_t, determining the appropriate lag order for $\Phi(L)$ and $\Theta(L)$ as well as estimating the parameters of $\Phi(L)$, $\Theta(L)$, and d (Peters, 1994, p.196).

1. Estimate d in the ARIMA(0,d,0) model $Y_t(1-L)^d = a_t$.

2. Define $u_t = Y_t(1-L)^d$.

3. Using the Box-Jenkins modeling procedure, identify and estimate the Φ and Θ in the FARIMA(p,0,q) model $\Phi(L)u_t = \Theta(L)a_t$.

4. Define $X_t = (\Theta(L))^{-1} * (\Phi(L)Y_t)$.
5. Estimate d in the FARIMA(0,d,0) model $(1-L)^d X_t = a_t$.
6. Check for the convergence of the d, Φ, and Θ; if not convergent, then goto step 2.

10.7 The Stable FARIMA

The stable fractional ARIMA (Kokoszka and Taqqu, 1996; Samorodnitsky and Taqqu, 1994) time series is defined by:

$$\Phi(L)X_t = \Theta(1-L)^{-d}u_t, \tag{10.21}$$

where the innovations u_t are i.i.d. and have infinite variance, and where d is a positive fractional number. The formula representation of the stable FARIMA(p,d,q) is basically identical to the Gaussian. The definition of $(1-L)^d$ is the same as for Gaussian FARIMA.

The difficulty for the stable case is the estimation of d, together with the parameters of the polynomials Φ and Θ. This problem will be treated in chapter 8.

Assuming the u_t to be i.i.d. with zero mean and following a stable law with $1 < \alpha < 2$, there is a unique moving average representation

$$X_t = \sum_{j=0}^{\infty} c_j u_{t-j}, \tag{10.22}$$

where $\sum_{j=0}^{\infty} c_j z^j = \frac{\Theta(z)}{\Phi(z)(1-z)^d}$, $|z| < 1$. Moreover, it is required that Φ and Θ have no zeros in common and no zeros in the closed unit disk $D = z : |z| \leq 1$. d is restricted to $d < 1 - 1/\alpha$.

The system has a unique solution if and only if $\Phi(z)$ has no roots in the closed unit disk $D = z : |z| \leq 1$. The sequence $X_t, n \in Z$ is then stationary and α-stable.

If $\Theta(z)$ has no unit roots in the closed unit disk $z : |z| \leq 1$ then the process is invertible. Invertibility is defined as $\sum_{j=0}^{\infty} c_j X_{t-j} = u_t$ and is very useful as it allows X_t to be expressed by previous observations.

Fractional stable noise. Assuming the innovations u_t of the process follow a stable law, $u_t \sim S_\alpha(\beta, \gamma, \mu)$, in the following the properties of fractional stable noise

FARIMA(0,d,0) are described by Samorodnitsky and Taqqu (1994, pp. 380). The stability index is restricted to $1 < \alpha < 2$:

FARIMA(0,d,0) is defined for all $d < 1 - 1/\alpha$. On the other hand, fractional stable noise is defined only for $0 < H < 1$. The condition $H < 1$ is equivalent to $d < 1 - 1/\alpha$. The lower bound, $H > 0$, expressed by $d = H - 1/\alpha$, becomes $d > -1/\alpha$. Thus, $0 < H < 1$ becomes $-1/\alpha < d < 1 - 1/\alpha$. This is an important range for fractional ARIMA.

10.8 The Multivariate FARIMA Model

$\tilde{X}_t$ is a vector of fractional integrated ARIMA processes, defined by

$$\Phi(L)D\left[(1-L)^d\right](\tilde{X}_t - \mu) = \Theta(L)a_t, \tag{10.23}$$

where L lag operator, $\Theta(L)$ and $\Phi(L)$ are matrix polynomials. $D\left[(1-L)^d\right]$ is a $k \times k$ diagonal matrix, with the values on the diagonal, and $d = (d_1, d_2, ..., d_k)$. The innovations a_t are a k-dimensional vector. The stationarity and invertibility conditions are the same as for the univariate case.

The stationary and invertible process has also an infinite moving average representation

$$\tilde{X}_t - \mu = \sum_{j=0}^{\infty} \Psi_j a_{t-j} \tag{10.24}$$

As in the univariate case, it can be proven that for the Vector-FARIMA model the cumulative impulse response $\sum_{k=0}^{j} \Psi_k$ diverges hyperbolically for $j \to \infty$ while a stationary and invertible vector ARMA model has geometrically convergent cumulative impulse responses.

10.9 Conditional Volatility And FARIMA

Andersen, Bollerslev, Diebold and Labys (1999) assume the log volatility of market returns to fall within a standard Gaussian ARMA class of models. They apply the natural logaritm to the volatility in order to reduce the skewness to the right of the original conditional volatility's distribution. By taking the log volatilities, the innovations of the ARMA are rendered more symmetric. In addition, this ensures that the volatility cannot become negative.

They furthermore realize the long-memory characteristics of the historical volatility series. This switches the initially proposed ARMA model into a FARIMA model.

However, it has to be noted that their proposed FARIMA model is different from the FIGARCH model applied by Bollerslev, Baillie, Mikkelsen (1996) which can also be written as a FARIMA model.

The steps of Andersen, Bollerslev, Diebold, and Laby (1999) to estimate the parameters of their model are organized as follows:

1. First they determine the degree of fractional integration d in the realized series.
2. Thus, their model is set up as follows: $u_t = (1-L)^d log(\sigma_t^2)$. This procedure should remove the LRD in the observed time series. The u_t are supposed to be i.i.d. normally distributed.
3. The u_t are analyzed if they are really i.i.d. by studying their SACF and SPACF.
4. The FARIMA model can now be applied for prediction.

As known, when applying stable non-Gaussian distributed random variables for the innovations of the credit model, $\epsilon_{i,t}$, a variance does not exist. Furthermore, conditional variance is not directly observable anyway.

The idea for the credit return model is to describe the absolute values of $\epsilon_{i,t}$ with a multivariate FARIMA(p,d,q) model based on stable non-Gaussian distributions. The next section develops a tractable multivariate model.

10.10 Developing A Multivariate LRD Process For The Innovations Of The Credit Return Model

After the theoretical introduction of the different long-memory models, the goal is to develop a tractable model on the basis of multivariate FARIMA that describes the behavior of the $\epsilon_{i,t}$ of the credit return model. Before stepping into the process of model building, the implications from the empirical examinations and introduced types of LRD processes are summarized first so far:

- Multivariate FIGARCH models (with / without fractional cointegration) turn out to be highly complex and are at most tractable for the Gaussian bivariate case.

- Multivariate FARIMA models are better tractable. Andersen, Bollerslev, Diebold and Labys (1999) have applied such a model for conditional log volatilities. As in the stable non-Gaussian case conditional variance does not exist, the forecasting of the absolute values of the $\epsilon_{i,t}$ is preferrable. The application of the MRS statistic has proven *highly significant LRD in the absolute values* of $\epsilon_{i,t}$.

One the one hand, the model should capture most of the predictive information and, on the other hand, be parsimonious and tractable in a multivariate framework. It should be easily integrated into the cointegrated VAR credit return model.

Furthermore, the model should keep the existing I(0)-I(1) cointegrating relations instead of designing a complicated fractional cointegration relation. Fractional integration should only be modeled as part of the $\epsilon_{i,t}$.

In addition, each fractionally integrated variable of the multivariate system should be allowed to have its own fractional differencing parameter. Each combination of credit rating/maturity can have its own d. (Such would not be possible with an approach of fractional cointegration that requires identical fractional differencing parameters for all included variables)

It is furthermore desirable to have a *robust and tractable estimator* for the fractional differencing parameter d, especially when dealing with heavy-tailed data ($\alpha < 2$). The applied estimator will be treated in chapter 11.

All these objectives mentioned can be achieved by a multivariate FARIMA model describing the absolute values of the stable distributed $\epsilon_{i,t}$. The expected skewness in the noise of the FARIMA model can be handled by stable distributions.

It has to be noted here that if long memory does exist in the absolute returns, then of course this information is also present in the signed returns themselves. The signed returns contain all information that can be obtained from the absolute returns series as well. However, there is not yet a measure defined that can describe the effects of LRD in the volatility directly for the signed returns.

The absolute returns as a measure for conditional volatility can be directly observed (unlike conditional variance that cannot be directly observed and does not exist for $\alpha < 2$). Forecasting the absolute values of the credit model's residuals means a direct prediction of the size of the residuals for the next period while the forecasting of conditional variance basically means a prediction of the probability distribution of the residuals.

The specification of the proposed multivariate FARIMA(p,d,q) for the credit return model is:

$$(|\epsilon_t| - \mu)(1 - L)^d \Phi(L) = \Theta(L) u_t. \tag{10.25}$$

ϵ_t is the five-dimensional vector of innovations for the credit return model with given maturity. u_t represents the five-dimensional vector of the noise of the FARIMA model. The other parameters are known from equation (10.23). Both the elements of $|\epsilon_t|$ and the elements of u_t are assumed to follow a stable law. The elements of u_t are dependent.

10.11 Conclusion

This chapter can be summed up as follows:

- The theory of fractional integration and long memory has been explained.
- The innovations of the credit return model have been analyzed for LRD.
- An overview of various fractionally integrated processes has been presented. Moreover, an appropriate choice to describe the long memory in the credit return model has been made.

The important findings of this chapter are:

- The LRD in the signed innovations $\epsilon_{i,t}$ is only measurable for the speculative rating grades, but still relatively weak. Instead, LRD is highly significant in the absolute values of $\epsilon_{i,t}$.
- LRD is not found in the noise of the GARCH/EWMA model of chapter 9 for the credit returns.
- Multivariate fractional models for conditional volatility (e.g. FIGARCH) turn out to be highly complex in the five-dimensional case and are not tractable.
- It makes sense to keep the original I(1)-I(0) cointegrating relation instead of implementing fractional cointegration that requires identical values for d in each equation. Instead, the chosen multivariate stable FARIMA model allows an individual d for each equation.

Thus, equation (10.25) is a new and suitable multivariate approach to describe the LRD in the volatilities of the multivariate credit return model. It is tractable, bears flexibility, and does not require additional restrictions.

The following chapter introduces the chosen estimator for the fractional differencing parameter d: A suitable estimator for the given problem is a modification of the conditional sum of squares (CSS) estimator. The modification is necessary as this estimator is per se only applicable to Gaussian time series and cannot be used for stable distributed time series. A test for the robustness of the estimator is presented as well. The significance of the estimates is tested with a special moving-blocks bootstrapping procedure designed for long-memory samples.

Chapter 11

Estimation Of The Long-Memory Parameter For The Credit Return Model

This chapter describes the modification of the Conditional Sum of Squares (CSS) Estimator for the α stable case with $\alpha < 2$.

The chapter is organized as follows:

- The known types of estimators are briefly reviewed.
- The CSS estimator is introduced and its properties are discussed.
- Finally, section 11.2.2 demonstrates how to transform the original CSS estimator so that it can be used for the case $\alpha < 2$.

11.1 Review Of Existing Estimators

Several types of estimators for FARIMA models have been suggested in the literature over time. Geweke and Porter-Hudak (1983) developed a log periodogram estimator, based on the regression of ordinates of the periodogram of the time series on a trigonometric function. It was followed by an averaged periodogram estimator of Robinson (1994). Fox and Taqqu (1986) have proposed a maximum likelihood estimator in the frequency domain. This estimator assumes conditional normality. It was extended to the α-stable case by Kokoszka and Taqqu (1996). Sowell (1992) has developed a full maximum likelihood estimator (MLE) for normally distributed innovations. A disadvantage of the estimator is that it is computationally quite burdensome because of the inversion of the large autocorrelation matrix and the fact that each element is a nonlinear hypergeometric function. Chong (2000) proposed the estimation of the fractional differencing parameter via the partial autocorrelation function (PACF).

11.2 The Conditional Sum Of Squares Estimator

11.2.1 Properties Of The CSS Estimator

The alternative conditional sum of squares (CSS) estimator is computationally less demanding. It is also capable of dealing with quite complicated FARIMA models having non-normal disturbances[1] or exhibiting time-dependent heteroscedasticity. Hosking (1984) has suggested it for fractional processes first. The CSS estimator is asymptotically equivalent to full MLE provided that the initial observations, eg. X_0, X_{-1}, X_{-2} are kept fixed.

The (Gaussian) CSS estimator minimizes the sum

$$S(\lambda) = \frac{1}{2}log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{t=1}^{n} u_t^2, \tag{11.1}$$

where the u_t are the residuals of the model which are assumed to be Gaussian i.i.d.

CSS is advantageous for all processes where full MLE turns out to be intractable for FARIMA models. Chung and Baillie (1993) have done simulation studies about the performance of the CSS estimator. They were especially interested in the behavior of the estimator when the mean μ of the FARIMA process is unknown. μ can be estimated in advance either by (i) the sample mean or by (ii) the median. The median is more robust to outliers, especially in small samples. Alternatively, μ can be estimated simultaneously with d by (iii) the CSS estimation procedure. Significant biases towards d occur with simultaneous estimation of μ and d when the number of samples is low.

If d is estimated conditional on known μ, the magnitude of the bias drops. Furthermore, an increase of the sample size n also reduces the bias of the estimates of d. It has to be noted that the sample sizes Baillie and Chung have worked with are very small (100 - 300 samples). The time series applied for estimating the credit return model have more than 2000 samples, so there should be less bias. In order to reduce bias for $\hat{d}$, the estimation of d with CSS will be performed conditional on known μ - as μ is intended to be estimated in advance. Thus, the applied estimator is supposed to be robust.

A useful check of the appropriateness of the estimator for mean μ or a possible bias is to use all three estimation techniques (sample mean, sample median, CSS)

[1] e.g. the application of the student-t distribution, see Chung and Baillie (1993).

and compare the obtained results. Radical differences would indicate a specification problem.

Clearly, an estimator based on conditional sum of squares cannot be used for heavy-tailed time series ($\alpha < 2$) without modification. In the following, a proper modification is derived in order to make the CSS estimator applicable for heavy-tailed time series.

11.2.2 Modification of The CSS Estimator

The modification of the CSS estimator is actually done by transforming the time series the estimator is applied to. The known Gaussian CSS estimator is then used for the transformed time series.

First, the tail index α of the time series must be determined. For the time series of the residuals ϵ_t, the value of α will be between 1 and 2. In this case not the conditional sum of squares but the conditional sum of p-th powers has to be applied for proper estimation of the parameters, with $1 \leq p < \alpha$.

For estimation procedures based on least-squares techniques in the Gaussian domain, the p-th sample absolute moment $\hat{m}_p = 1/n \sum_{i=1}^{n} |\hat{u}_i|^p$ has to be applied in the stable non-Gaussian domain, where n represents the sample size and α is restricted to $1 \leq \alpha < 2$). The fact is proven in the following.

Lemma 1 *The u_t are i.i.d. disturbances*[2] *in the normal domain of attraction of a symmetric α-stable random variable X. Let X be a symmetric α-stable random variable with characteristic function $Ee^{i\theta X} = e^{|\delta\theta|^\alpha}$, $0 < \alpha < 2$.*

Then $E|X|^p = C(p, \alpha)\sigma^p$, $0 < p < \alpha$.
where $C(p, \alpha) = 2^{p+1}\Gamma(\frac{p+1}{2})\frac{\Gamma(\frac{-p}{\alpha})}{\alpha\sqrt{\Pi}\Gamma(-p/2)}$ and $\Gamma(.)$ is the Gamma function.

Therefore, for $0 < p < \alpha$, the estimator $\hat{\sigma}_p$ for σ is defined as $\hat{\sigma}_p = \left(\frac{\hat{m}_p}{C(p,\alpha)}\right)^{1/p}$,

where $\hat{m}_p$ is the p-th sample absolute moment $\hat{m}_p = \frac{1}{n}\sum_{i=1}^{n} |\hat{u}_i|^p$.

From this it can be concluded: $\hat{\sigma}_p \to \sigma$, almost surely, that is: $\hat{\sigma}_p$ is a strongly consistent moment estimator for σ, $0 < p < \alpha$. If u_t follows a stable law with index α, then all absolute moments $E(u_t)^p$ of order $0 < p < \alpha$ exist. For $p \geq \alpha$, the moments are infinite.

The proof is given in Rachev and Mittnik (2000).

[2]See Rachev and Mittnik (2000).

Next the limiting behavior of $|u_i|^{p/2}$ is analyzed for $0 < p < \alpha$.
The p-th absolute sample moment is written as

$$\frac{1}{n}\sum_{i=1}^{n}|u_i|^p, \tag{11.2}$$

where $0 \leq p < \alpha < 2$ and u is a symmetric α-stable random variable which has a zero mean and n is the sample size. Recall that, as $\lambda \to \infty$, $P(|u_i| > \lambda) \approx C\lambda^{-\alpha}$, and thus

$$P(|u_i|^p > \lambda) \approx \frac{C}{\lambda^{\alpha/p}} \text{ as } \lambda \to \infty. \tag{11.3}$$

and $p > \alpha/2$.

Thus, $|u_t|^p$ is in the domain of attraction of the stable Paretian law with index α/p.

Feller's *Central Limit Theorem* for the α-stable law implies

$$n^{(\alpha-p)/\alpha}\left(\frac{1}{n}\sum_{j=1}^{n}|u_i|^p - E(|u_i|^p)\right) \xrightarrow{\text{w}} S_{\alpha/p}\ , \tag{11.4}$$

where $S_{\alpha/p}$ is a stable random variable with index α/p (see Feller (1971)).

Since $1 \leq p < \alpha$, then the variance

$$var(|u_i|^{p/2}) = E(|u_i|^p) - \left(E|u_i|^{p/2}\right)^2 \tag{11.5}$$

is finite, and therefore, $|u_i|^{p/2}$ is in the domain of attraction of a Gaussian law:

$$n^{1/2}\left(1/n\sum_{i=1}^{n}|u_i|^{p/2} - E(|u_i|^{p/2}\right) \xrightarrow{\text{w}} N(0, var(|u_i|^{p/2})), \tag{11.6}$$

while the variance $var(|u_i|^{p/2})$ can be expressed by equation (11.5). Thus, if the time series of $|\epsilon_{i,t}|$ is transformed according to

$$z_{i,t} = |\epsilon_{i,t}|^{p_i/2}, \text{ with } i = 2...6, \tag{11.7}$$

the CSS estimator can be applied to it, where $1 \leq p_i \leq \alpha_i$.[3] Analyzing the tail indices α_i of the $\epsilon_{i,t}$, $1 < \alpha_i < 2$ will be found in all cases. Setting $p_i = 1$, the final estimator is suitable for all $1 < \alpha_i < 2$ and becomes equal for all time series $|\epsilon_{i,t}|$. Thus, it will be set for all equations of the credit return model: $z_{i,t} = |\epsilon_{i,t}|^{1/2}$. The CSS estimator can now be applied to the model $(z_{i,t} - \mu_{z,i})(1-L)^{d_i}\Phi_i(L) = \Theta_i(L)u_{i,t}$. Measuring the stability indices of the $z_{i,t}$ it can be found that those are close to 2 or even equal to 2.

The estimator will be applied to fit the stable FARIMA(0,d,0) model first:

$$(z_t - \mu_z)(1-L)^d = u_t. \tag{11.8}$$

Then it has to be tested, if AR or MA components are significant and have to be added.

Annotation: It does not matter if the random variable of interest with $\alpha_i < 2$ is transformed from $|\epsilon_{i,t}|$ into $|\epsilon_{i,t}|^{p_i/2}$ or if the estimator is modified from sum of squares to the sum of p-th powers. The resulting sums are identical.

The modified CSS estimator for stable distributed time series with $\alpha < 2$ is a new estimator to obtain the fractional differencing operator for heavy-tailed processes, eg. stable FARIMA. As demonstrated - the limiting behavior of the estimator is in the Gaussian domain of attraction.

In order to obtain the estimate for μ, the sample mean will be chosen. As mentioned, the sample size is relatively large and because of the transformation $|\epsilon_t|^{1/2}$ the impact of outliers is reduced. Therefore, the selection of the sample median as an estimator for the mean of the process should not be an improvement compared to the sample mean.

The next section discusses two testing methods to validate the estimator:

- First, it presents how to test the robustness of the modified CSS estimator for stable time series.
- Second, it describes a moving-blocks bootstrapping procedure that empirically derives the asymptotic distribution of the estimates and tests their significance.

11.3 Checking Inference For FARIMA

11.3.1 Robustness Of The Estimator

Assessing the robustness of the estimator, the stability of the estimates of the sum of p-th powers has to be examined for moving p. Assuming p would move from 1

[3] $i = 1...5$ represents the five credit rating grades AA, A, BBB, BB, B.

to α, how would this affect the estimates of the parameter d and eventual other parameters? Thus, the parameters have to be estimated for various values of p, $p \in [1; \alpha)$. Several values for p are selected from the interval in ascending order with constant distances between subsequent p.

A robust estimator should exhibit only slight changes for the values of the estimates over the range of p.

Parametric bootstrap methods show that inference in FARIMA processes is quite reliable in finite samples (Ooms and Doornik, 1999). The significance of the parameter estimates confirms the choice of the model. Especially, the determination of the significance of the fractional differencing parameter is difficult. To empirically obtain the asymptotic distribution of the estimate, traditional resampling methods do not work. Thus, a special bootstrapping method has been developed for this problem.

11.3.2 Significance Of The Estimates

In order to evaluate the significance of the estimates of the fractional differencing parameter d, a moving-blocks bootstrapping procedure will be applied. The bootstrapping procedure provides the empirical distribution of the estimate. For the fractional differencing parameter d, the variance of the estimate cannot be calculated directly (as for example in a linear regression). In order to test the hypothesis $H_0 : d = 0$ against $H_1 : d \neq 0$, the asymptotic distribution of the estimate $\hat{d}$ is required.

Traditional bootstrap and jackknife methods, invented by Efron (1979), are based on independent and identically distributed obeservations. However, this method is not suitable for a long-memory process as it destroys the long-range dependence in the sample.

To obtain a variance for the estimated long-memory parameter $\hat{d}$, a moving-block bootstrapping algorithm is suggested. This method has been introduced for cases when the data is not independent. Maharaj (1999) has proposed such a test for the estimation of the fractional differencing operator $\hat{d}$, determined on a stationary dependent data set.[4] The idea is to keep the dependence structure between the observations. So-called *moving blocks* are defined from the original time series of length n, $x_1, ..., x_n$. For the original time series, there exist $n - b + 1$ moving blocks consisting of b consecutive observations. These moving blocks are $B_1, ..., B_{n-b+1}$. The j-th block consisting of b consecutive observations is $B_j = \{x_j, ..., x_{j+b-1}\}$. A resampling of the blocks is performed by randomly drawing k blocks with replacement, and the blocks are pasted together forming a new time series of $k * b \approx n$ observations again. This produces a set of blocks $\{B_1^*, ..., B_k^*\}$. It

[4] In her publication the test is applied for a different type of estimator for d.

is important to select a block size b that is large enough to capture the long-range dependence among the observations (the autocorrelation structure of the samples should not be destroyed). It can be assumed that observations that are more than b units apart show only weak dependence or are almost independent. As mentioned, a resampling from independent observations clearly cannot be applied as it would of course destroy the dependencies.

The moving-blocks bootstrapping method is a non-parametric method and thus, does not require that a parametric or semi-parametric model has to be fitted to the dependent data first.

According to Hall, Horowitz and Jing (1995), the choice of the block size b depends on the context of the problem and is a data driven selection. Asymptotic theory makes the block size increase with the sample size. The bootstrap test statistic is a metric with $\hat{d}$ as the estimate of the original time series and d as the value of the fractional differencing operator in the null-hypothesis. The test statistic is

$$W = |\hat{d} - d|. \tag{11.9}$$

To obtain a first order asymptotic approximation to the distribution of W under the null hypothesis, the bootstrap method is performed J times. Each bootstrap sample delivers an estimate $\hat{d}^*$, and the test statistic W^* is calculated for each: $W^* = |\hat{d}^* - \hat{d}|$. All J values of W^* form the bootstrap estimate of the distribution of W. The bootstrap p^*-value is an estimate of the p-value that would be associated with the test statistics W. p^* is obtained as the frequency that the value of W^* exceeds W, divided by J:

$$p^* = \#(W^* > W)/J. \tag{11.10}$$

Introducing a nominal significance level δ, H_0 is rejected if $p^* < \delta$.

The moving-blocks bootstrap procedure is applied to each estimate of the altogether 15 fractional differencing parameters d_i in the above introduced multivariate model. The time series $z_{i,t}$ are divided into 10 blocks, each having a length of 217 units. This indicates a ratio between block size b and sample size T of $b = T^{0.7}$ which is similar to the example in Maharaj (1999). J is set to 1000, meaning that each time series is resampled 1000 times. The block length (217 units) is sufficient not to destroy the long-memory structure in local areas of the time series. It is furthermore sufficient to detect if the long-memory parameters strongly vary from block to block, and if for certain resamplings the long memory would vanish. For the estimation of the d_i, the infinite lag polynom was cut at lag 500.

Analyzing the obtained empirical distribution of W, it is important to check if it is moving towards a normal distribution. If this can be confirmed, the p-values of the hypothesis test can be calculated.

The introduced moving-block bootstrapping test applied to the CSS estimates helps to reveal if the processes governing different blocks of a time series are homogeneous or, if they are inhomogeneous and the obtained estimate for the fractional differencing turns out not to be significant.

Berkowitz, Birgean and Kilian (1999) however point out that there could be a bias in the estimate of a statistic when resampling blocks of b consecutive observations from the original sample. The bootstrap procedure treats consecutive blocks as conditionally independent. The jumps between the blocks are reported as the reason for the bias of the estimates.

Indeed, it might be questionable to what degree such jumps are quantitatively relevant within large sample sizes of more than 2000 observations as in this case here. Thus, the given moving-blocks bootstrapping specification complies with the underlying problem as autocorrelations decay hyperbolically over the lags. Only those observations that are very close to the left boundaries of the blocks face the impact of the boundaries. Due to the large size of the blocks, the fraction of affected observations should not be meaningful.

However, for the objective of the bootstrapping procedure is not the exact estimation of the desired parameter with the least bias. Instead, it is important to have a sufficient precision for testing the significance of the obtained parameter estimates $\hat{d}$.

11.4 Conclusion

The results of this chapter can be summed up as follows:

- A new estimator to obtain the fractional differencing parameter d for heavy-tailed times series has been developed, based on the modification of the Gaussian Conditional Sum of Squares (CSS) estimator. The advantage of this estimator is that it is less complex and difficult to implement than other estimators under the stable assumption. Additionally, it is less computationally burdensome than other estimators.

- The estimator is defined for p with $1 \leq p < \alpha$. The robustness of the estimator has to be tested for different values of p. The significance of the estimates of the fractional differencing parameter d can be tested by a moving-blocks bootstrapping procedure.

The following chapter presents the estimation results of the multivariate FARIMA for the credit return model. In addition, the significance of the estimates is tested with the introduced moving-blocks bootstrapping procedure. The accuracy of the long-memory model for Value at Risk forecasts is evaluated and compared with the forecast accuracy of the EWMA model of chapter 9.

Chapter 12

Empirical Long-Memory Analysis

This chapter presents the fitting results of the multivariate FARIMA model for the credit return model. The exact FARIMA specification is determined and the fractional differencing parameter d is estimated for each combination of credit rating / maturity. FARIMA(0,d,0) turns out to be the proper specification and the estimates of d are all significant.

Analyzing the accuracy in one-step-ahead Value at Risk forecasts, it is found that the long-memory model exhibits on average a better performance over all credit rating / maturity combinations than the formerly chosen EWMA model.

Furthermore, the chapter introduces copulas, which are a better and more natural description of dependence than Gaussian correlations. It can be demonstrated that the developed long-memory FARIMA model allows the application of a Gaussian copula, which is especially tractable for the high-dimensional case, for a random vector with stable marginals.

12.1 The Data

For building the cointegrated VAR with an LRD model in the residuals, a longer sample time series was used (2170 observations) compared to the VAR with traditional GARCH models. The estimation of LRD parameters usually require a larger set of data to obtain significant estimates. The parameters of short-memory GARCH processes are - which is known from experience - less stable over time and have to be updated regulary.

For modeling the transformed absolute values $|\epsilon_{i,t}|^{p/2}$, $1 \leq p < \alpha_i$, *the parameter* p *is set to* 1. This choice satisfies the condition $1 \leq p < \alpha_i$ for all i and enables an equal parameter p for all variables $|\epsilon_{i,t}|^{p/2}$.

12.2 Analysis Of The SACF

First, a graphical examination of the five residuals of each maturity (after regression on the common credit risk factor res_c) is performed.

Clustering volatility is both a sign of long memory and traditional autoregressive heteroscedatic models. The autocorrelations of LRD models show a hyperbolic decay. The following figure 12.1 presents the sample autocorrelation functions (SACF) for the transformed absolute values $|\epsilon_{i,t}|^{1/2}$ of the residuals of the 2, 10, and 30 - year-maturity cointegrated VAR.

As it can be seen from figure 12.1, all graphs show mainly positive, slowly decaying, autocorrelations. Slight negative autocorrelations are only measured at lags beyond 40. Especially, the B ratings exhibit a steady hyperbolic decay which indicates that they show a stronger long memory than the AA rated, and are therefore supposed to have a greater fractional differencing parameter d.

12.3 Estimation And Model Testing

The building and estimating of the proper FARIMA(p,d,q) specification for the residuals of each equation of the cointegrated VAR is performed step by step with validation following.

First, the time series $z_t = |\epsilon_t|^{1/2}$ are built. The first model to be estimated is the FARIMA(0,d,0):

$$(z_t - \mu_z)(1 - L)^d = u_t, \tag{12.1}$$

where

$$z_t = |\epsilon_t|^{1/2}. \tag{12.2}$$

The elements of vector u_t are supposed to be i.i.d. and follow a stable law. d is a vector containing the d_i for each maturity / credit rating combination.

If significant short-term autocorrelations would be found in the u_t, this would indicate a specification error. Thus, both the SACF and SPACF will be plotted and analyzed for the elements of u_t.

As the autocorrelations of z_t for the first lags are relatively low and do not decline sharply, the presence of a short-term AR polynomial can probably be denied. To demonstrate that there is no short term AR component, FARIMA(0,d,0) and FARIMA(1,d,0) are estimated simultaneously. The addition of a short-term AR component can be tested with the residuals-based F-test. The results of the estimation procedure are presented in table 12.1:

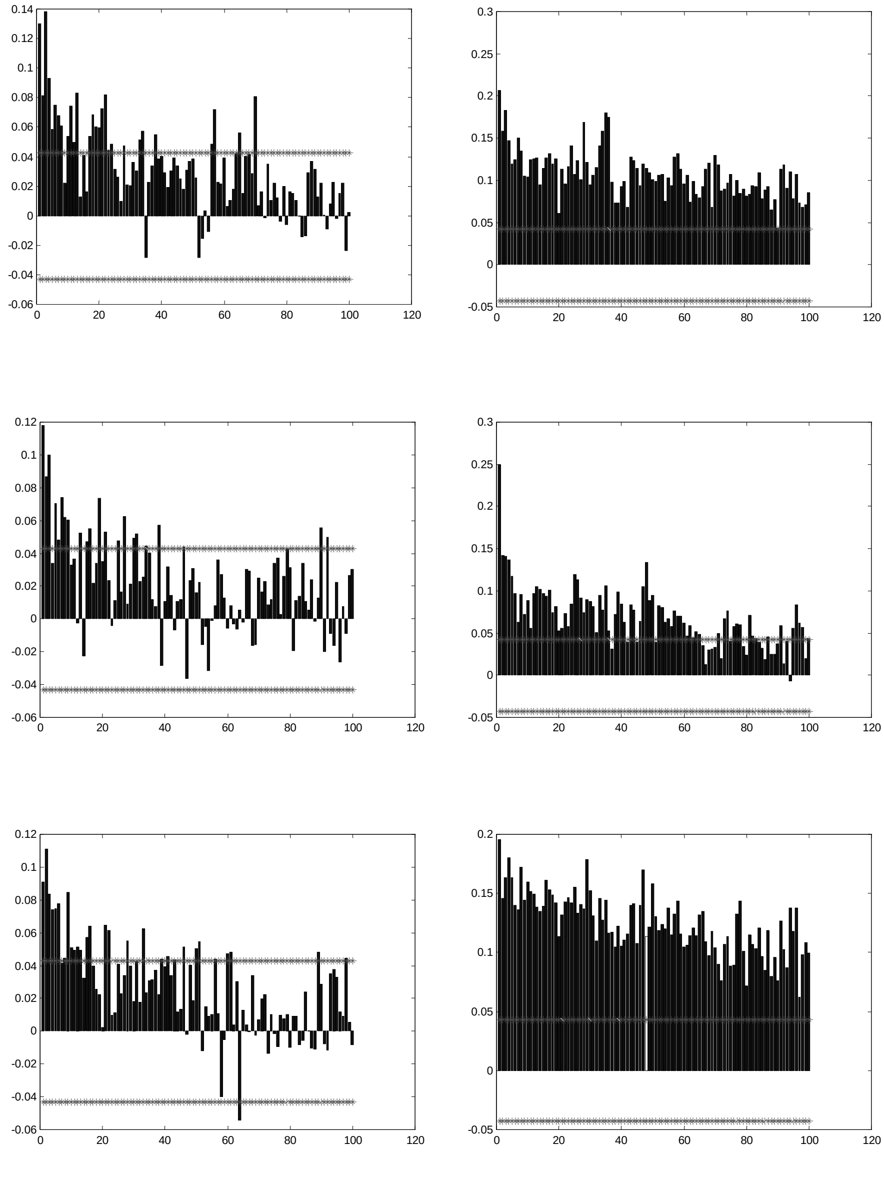

Figure 12.1: SACF of $|\epsilon|^{1/2}$. The plots for AA are on the left and the plots for B on the right, with maturities 2, 10, and 30 years (starting with upper row).

	$\hat{d}$	$\hat{d}$ AR(1)	AR(1)	p-Value F-Test
AA 2y	0.1337	0.0919	0.1082	0.1092
A 2y	0.1473	0.1020	0.1163	0.0807
BBB 2y	0.1411	0.1426	-0.0038	0.9540
BB 2y	0.1539	0.1503	0.0113	0.8356
B 2y	0.1757	0.1502	0.0746	0.1724
AA 10y	0.1320	0.1232	0.0219	0.7434
A 10y	0.1543	0.1760	-0.0512	0.4325
BBB 10y	0.1220	0.1529	-0.0772	0.2469
BB 10y	0.1559	0.1556	0.0007	0.9893
B 10y	0.1609	0.1671	-0.0180	0.7475
AA 30y	0.1159	0.0961	0.0559	0.3867
A 30y	0.1084	0.1067	0.0043	0.9488
BBB 30y	0.1212	0.1306	-0.0241	0.7192
BB 30y	0.1595	0.1669	-0.0227	0.6704
B 30y	0.1763	0.1782	-0.0053	0.9173

Table 12.1: CSS-estimates of d for FARIMA(0,d,0) and FARIMA(1,d,0), the parameter value for the AR(1) component and the p-value of the F-test.

The second column from the left is the CSS-estimate of d for the FARIMA(0,d,0) model. The $\hat{d}_i$ lie all in the interval $[0.1084, 0.1763]$. The fractional differencing parameters d_i seem to *increase from higher to lower credit quality.* The addition of an AR(1)-component does not improve any of the 15 models, only the A-2y-equation has a probability value of less than 0.10 for the F-test.

The validity of the model specification is further examined by analyzing the behavior of the $u_{i,t}$. The plots of SACF and SPACF for all equations indicate no short-term AR or MA component, it can be confirmed that the $u_{i,t}$ are i.i.d.

In figure 12.2, the kernel plots of the empirical density of the elements of u_t show that the distribution is extremely skewed to the right. This is confirmed by the results of fitting the empirical distributions with the four-parameter stable distribution.

The shape of the distributions does not exhibit strong peakedness and the stability index mainly appears to be close to the Gaussian case. This is related to the transformation $|\epsilon_t|^{1/2}$ which causes a contraction of the range of the sample data and especially outlier-observations are shifted closer to the center. In case of $\alpha_i < 2$, the skewness property is expressed by a skewness parameter $\beta_i \neq 0$. However, if $\alpha_i = 2$ is obtained by the ML-procedure, β_i is fixed to be zero.

The parameters of the stable fitted u_t are given in table 12.2.

The α_i of the fitted u_t vary between 1.7 to 2. It has to be noted that the

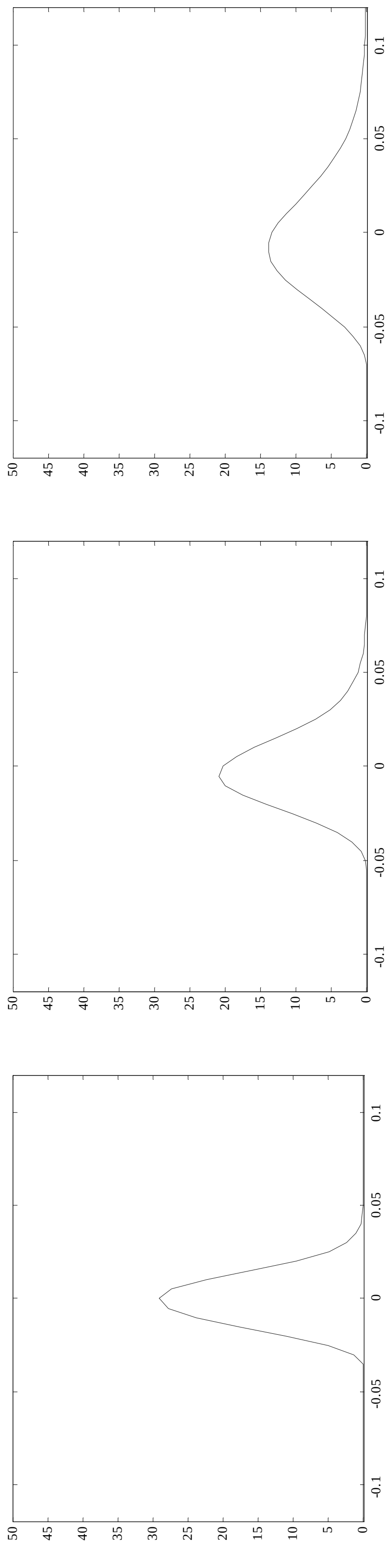

Figure 12.2: Kernel plots of the probability density for selected elements of u_t: the 2-year (left), 10-year (center), and 30-year (right) of BBB.

	α	β	γ	δ
2y AA	2.0000	0	0.005656	-0.0002655
2y A	1.718539	1	0.004902212	-0.001619482
2y BBB	2.0000	0	0.005721	-0.0006651
2y BB	1.6929973	1	0.006014521	-0.002246872
2y B	1.7472074	1	0.006706087	-0.00201588
10y AA	1.7916044	1	0.010917696	-0.003059942
10y A	2.0000	0	0.01232	-0.001362
10y BBB	1.7214655	1	0.010692806	-0.00403184
10y BB	1.7505914	1	0.013735287	-0.004745075
10y B	1.6851007	1	0.01536055	-0.00613727
30y AA	1.7641642	1	0.017257767	-0.004701462
30y A	2.0000	0	0.02039	-0.0006025
30y BBB	1.7150517	1	0.017816765	-0.005805664
30y BB	1.6855322	1	0.022266536	-0.007859206
30y B	1.7011069	1	0.025222148	-0.008027554

Table 12.2: Stable fitting of u_t allowing individual stability indices. Some elements exhibit $\alpha_i = 2$.

skewness parameter β_i is equal to 1 for all cases with $\alpha_i < 2$. In case the fitting procedure determines $\alpha_i = 2$ the skewness property cannot be modeled.

In order to keep the extreme skewness property which is important for the accuracy of Value at Risk measurement, the skewness parameter β can be fixed in advance. Thus, the fitting of the elements of u_t by Maximum Likelihood is now performed with 3 parameters only, setting $\beta = 1$. This has an impact on the fitted parameters of the elements 2y AA, 2y BBB, 10y A, and 30y A. The results are given in table 12.3.

As the u_t exhibit strong skewness to the right, the right tail is important for the Value at Risk estimation. The results of the four-parameter fit is that for those elements of u_t which have $\alpha < 2$, the skewness parameter becomes 1. For those elements of u_t with $\alpha_i = 2$ originally, setting $\beta_i = 1$ results in α_i around 1.7 which is similar to the α_i of the other elements of u_t. By doing this the skewness of all elements of u_t is kept.

In order to model the dependence between the elements of u_t applying the sub-Gaussian approach[1], a common index of stability for all equations would be required. The average of all stability indices is 1.8690. As this value is less than 2, the extreme skewness property of the u_t is kept as β is allowed to be different from zero. The approach of Rachev, Khindanova and Schwartz (2001) splits each

[1]See Rachev, Schwartz and Khindanova (2001)

	α	γ	δ
2y AA	1.7532765	0.00512446	0.000648712
2y A	1.7193429	0.004883423	0.000678346
2y BBB	1.7332084	0.005161207	0.000714781
2y BB	1.6910177	0.006007861	0.000938096
2y B	1.7433563	0.006655177	0.000801312
10 y AA	1.793435	0.010933613	0.000660316
10 y A	1.8361544	0.011384275	0.000528621
10 y BBB	1.7212313	0.01065713	0.000948069
10 y BB	1.7512531	0.013690179	0.000877043
10 y B	1.6871254	0.015316269	0.002055915
30 y AA	1.7644	0.017209445	0.001980125
30 y A	1.7448896	0.017081686	0.002081445
30 y BBB	1.717224	0.017819211	0.002710935
30 y BB	1.6849063	0.022264373	0.004202894
30 y B	1.7032394	0.025113989	0.004566008

Table 12.3: Stable fitting of u_t under the restriction $\beta = 1$, allowing individual stability indices.

element of u_t into a dependent symmetric and into an independent skewed part. However, the sub-Gaussian approach exhibits some disadvantages:

- The necessary choice of a common stability index for all elements of the vector u_t results in a loss of goodness of fit.
- In addition, the estimation of the Gaussian correlation between two stable distributed time series poses the problem of proper truncation of the time series.
- Furthermore, it is doubtful that the dependence between two random variables is represented appropriately by a linear measure such as correlation.

In order to overcome the above mentioned deficiencies of the sub-Gaussian approach, advanced approaches use *copulas* to model the dependence between random variables. Copulas are a *natural* way to model dependence. The type of copula that will be chosen for a skewed stable random vector here does neither require common stability indices nor splitting each element of the random vector into a skewed independent and a symmetric dependent component. The approach is capable of describing the dependence structure of a high dimensional multivariate stable random vector. This should result in a more accurate Value at Risk as each element of u_t is allowed to keep its original stability index which determines the density in the tail.

Copulas are introduced in section 12.7.

12.4 Results Of Robustness Check

The robustness of the estimator for the fractional differencing parameter d has to checked depending on the value of p. Transforming the vector of interest $|\epsilon_t|$ by taking the $p/2$-th power , $1 \leq p < \alpha$, it has to be checked how the estimates of d behave for a range of values taken from $p \in [1, \alpha)$.

$\hat{d}$	1.00	1.05	1.10	1.15	1.20	1.25	1.30	1.35	1.40	1.45
AA 10y	0.1569	0.1578	0.1586	0.1593	0.1599	0.1604	0.1607	0.1610	0.1611	0.1611
B 10y	0.1631	0.1643	0.1655	0.1666	0.1677	0.1687	-	-	-	-

Table 12.4: Testing the robustness of the estimator for d, depending on the value of p.

The robustness of the estimator is tested for two residual series $\epsilon_{i,t}$; one is the AA with 10-year maturity, the other is the B with 10-year maturity. The two series have been chosen as representatives of the lowest and the highest credit quality within the model. The α_i for 10-year AA residuals is 1.47 while for 10-year B residuals it is 1.27.

As it can be seen from figure 12.3, the obtained estimates for various p differ only slightly. Within the range of $p \in [1, \alpha)$, the estimate of d_i for 10-year AA increases by 2.68% and for 10-year B by 3.43%. Thus, it can be concluded that the chosen CSS-estimator for d is quite robust. As the estimates remain quite stable for different values of p, it is justified to set $p = 1$ for the estimator in each of the series $|\epsilon_{i,t}|^{p/2}$. The results of table 12.4 indicate that the robustness of the estimator does not seem to be influenced by the credit quality. It is reasonable to assume the same for the series of the 2-year and 30-year maturities as well.

12.5 Results Of Moving-Block Bootstrapping

The results of the moving-blocks bootstrapping procedure are given in tables 12.5 - 12.7. For each time series $|\epsilon_{i,t}|^{1/2}$, the fractional differencing parameter has been estimated with 1000 resamplings. It is desirable that the density of the obtained empirical distribution of the estimates moves towards a normal distribution. The skewness and the kurtosis of the obtained empirical distributions will be measured and compared with the values that are characteristic for the Gaussian. Possible biases are usually indicated by skewed distributed empirical results.

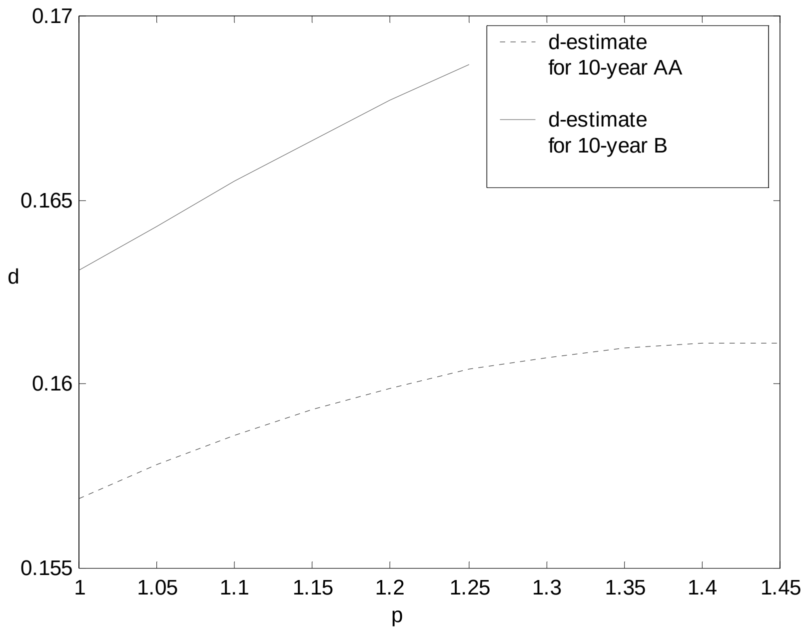

Figure 12.3: Robustness of CSS-estimator for various values of p.

2 year	AA	A	BBB	BB	B
estimate (1)	0.133656	0.147312	0.141121	0.153875	0.175750
bootstrap-mean (2)	0.127999	0.132907	0.135833	0.140305	0.159345
bootstrap-std	0.028362	0.027818	0.020811	0.021837	0.025927
diff. (1) - (2)	0.005656	0.014405	0.005287	0.013569	0.016404
skewness	-0.326504	0.017937	-0.769932	-0.976921	-0.054311
kurtosis	3.034757	2.531610	3.996542	4.509883	2.566734

Table 12.5: Analysis of the empirical distribution of moving-block bootstrapping estimates $\hat{d}$, for 2-year maturity.

The first line (”estimate (1)”) in tables 12.5 - 12.7 provides the estimates for the original time series. The second line gives the mean of the moving-blocks bootstrapping estimates (”bootstrap-mean (2)”). The difference between the estimate and the bootstrap-mean is given by ”diff. (1) - (2)”. ”bootstrap-std” is the standard deviation of the empirical distribution of bootstrap estimates. The deviation from the Gaussian distribution is measured by the empirical kurtosis and skewness. The skewness of the distribution of the bootstrap estimates $\hat{d}_i$ is computed as

10 year	AA	A	BBB	BB	B
estimate (1)	0.131980	0.154250	0.121976	0.155875	0.160875
bootstrap-mean (2)	0.123733	0.133348	0.106432	0.206238	0.201524
bootstrap-std	0.031002	0.034637	0.037061	0.065315	0.049112
diff. (1) - (2)	0.008247	0.020901	0.015544	-0.050363	-0.040649
skewness	-0.291501	0.000630	0.003259	0.469126	0.122658
kurtosis	2.744693	2.919323	2.726910	2.355432	2.806747

Table 12.6: Analysis of the empirical distribution of moving-block bootstrapping estimates $\hat{d}$, for 10-year maturity.

30 year	AA	A	BBB	BB	B
estimate (1)	0.115882	0.108359	0.121164	0.159500	0.176343
bootstrap-mean (2)	0.111381	0.099378	0.110523	0.134360	0.163698
bootstrap-std	0.018348	0.013727	0.029297	0.033559	0.033403
diff. (1) - (2)	0.004500	0.008980	0.010640	0.025140	0.012644
skewness	-0.389375	-0.208527	-1.416338	-1.857207	-0.267275
kurtosis	3.340196	2.936556	5.600807	7.769564	2.538950

Table 12.7: Analysis of the empirical distribution of moving-block bootstrapping estimates $\hat{d}$, for 30-year maturity.

$$S = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{d}_i - \bar{d}}{\hat{\sigma}} \right)^3 , \qquad (12.3)$$

where $\hat{\sigma}$ is based on the biased estimator for the variance. Positive skewness means a long right tail and negative skewness a long left tail. For the normal and all other symmetric distributions, the skewness is zero. N is the number of samples (i.e. the number of resamplings with the bootstrap procedure). The kurtosis is defined as

$$K = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{d}_i - \bar{d}}{\hat{\sigma}} \right)^4 , \qquad (12.4)$$

where $\hat{\sigma}$ is based on the biased estimator for the variance. For the normal distribution, the kurtosis is 3. For $K > 3$, the distribution is peaked, for $K < 3$ the distribution is flat. N denotes the number of samples.

In general, the means of the bootstrap estimates deviate only slightly from the estimates of the original series. Exceptions are the 10-year BB and B models. However, the differences between original estimates and bootstrap-mean are always less than one standard deviation of the bootstrapping estimates. For the speculative grades, slight skewness can be observed for the distribution of the estimates which means a minor deviation from the shape of the normal distribution. The estimates for the 30-year BBB and BB exhibit the largest values for the empirical kurtosis and a distinctive skewness. While for the investment grade series the bootstrap estimates are almost normally distributed, the estimates for BB and B are slightly skewed to the left. I remains a question if this could be overcome by choosing a larger number of bootstrapping samples. As the speculative grades tend to have a larger d, the autocorrelations for longer lags are larger, and thus, could lead to greater jumps between consecutive blocks. Increasing the length of the blocks might be a remedy. On the other hand, the 10-year A investment grade also exhibits a relatively large $\hat{d}$ of 0.15425, but skewness of the empirical distribution of the estimate is obviously not the case. Generally speaking, the empirical asymptotic distributions of the bootstrapping estimates for the fractional differencing parameter are roughly normal with few exceptions. Performing a larger number of resamplings should improve the shape of the distribution towards normal. After calculating $W = |\hat{d}^* - \hat{d}|$ in order to obtain W^*, the resulting bootstrap-based p-values are given in table 12.8.

p-values	AA	A	BBB	BB	B
2 year	0	0	0	0	0
10 year	0	0	0.0010	0.0680	0.0060
30 year	0	0	0	0.0050	0

Table 12.8: Bootstrap-based p-values for estimates of d.

Except for the estimate in the 10-year BB model, all estimates of parameter d are significant at a level of 0.01. This justifies both the inference of a fractionally integrated process for $|\epsilon_{i,t}|^{1/2}$ and the selection of the FARIMA(0,d,0) model. Additionally, it confirms the performance of the chosen estimator for d.

The following section 12.6 examines the dependence among the elements of u_t, both for the centers of their distributions and for the tails. In section 12.7 a copula will be introduced that is capable of describing the dependence between skewed stable distributed elements of high-dimensional multivariate stable random vectors.

12.6 Analyzing Dependence In u_t

After the FARIMA process has been estimated and the significance of the estimates has been demonstrated, the vector u_t representing the FARIMA-noise is examined.

Analyzing the dependence between the elements of the u_t, correlations are measured for the whole sample without truncation first. The results are presented in table 12.9.

2 year	AA	A	BBB	BB	B
AA	1	0.4212	0.3902	0.2211	0.1628
A	0.4212	1	0.4222	0.24	0.1478
BBB	0.3902	0.4222	1	0.2411	0.1453
BB	0.2211	0.24	0.2411	1	0.2986
B	0.1628	0.1478	0.1453	0.2986	1

10 year	AA	A	BBB	BB	B
AA	1	0.5603	0.5255	0.2765	0.2189
A	0.5603	1	0.5011	0.2538	0.2043
BBB	0.5255	0.5011	1	0.2779	0.1966
BB	0.2765	0.2538	0.2779	1	0.3822
B	0.2189	0.2043	0.1966	0.3822	1

30 year	AA	A	BBB	BB	B
AA	1	0.3531	0.3649	0.154	0.1635
A	0.3531	1	0.3584	0.1728	0.1861
BBB	0.3649	0.3584	1	0.1475	0.1393
BB	0.154	0.1728	0.1475	1	0.3347
B	0.1635	0.1861	0.1393	0.3347	1

Table 12.9: Correlations between the elements of u_t, taking the whole sample without truncation.

Examining the un-truncated samples, it is found that they exhibit considerable correlations. However, it has to be analyzed if those correlations are mainly present due to the tails of the right-skewed distributions of u_t or if they are equally present both in the center and tail. To get a deeper insight, the correlations are now calculated separately for the observations in the tail and for the observations in the center.

The distinction if an observation belongs to the center or the tail is made according to the following rule: An observation (vector) at time t, u_t, is counted as an observation of the tail if at least one of its elements is greater than the 90% quantile of the empirical distribution. All other observations are counted for the center. This means, that for the 2-year maturity, 1144 vector-observations lie in the center and 535 are in the tails. The 10-year maturity has 1154 in the center and 525 in the tails, the 30-year maturity has 1107 in the center and 572 in the tails. The sample size of $\hat{u}_t$ is 1679 as the residuals $\hat{u}_t$ cannot be calculated for the first 500 observations.[2]

The observations in the tails usually have a greater distance from the series's mean than the observations in the center. Thus, observations in the tail areas cause a larger contribution to the correlation coefficient than an observation taken from the center. However, the *kind of conditioning* for the tail is important. Therefore, at this stage, a closer look at the problem of tail correlations is required.

Having a sample of observations with a given threshold, the tail of a distribution can be perfectly described by the three parameters: the tail probability, the dispersion parameter, and the tail-index.

In the multivariate case, asymptotic independence of extreme returns is reached very often. The following rule is presented by Longin and Solnik (2001): If all correlation coefficients between any two components of a multivariate normal process are different from $+/-1$, then the return exceedances tend to independence as the thresholds used to define the tails tend to the upper endpoint of the distribution ($+\infty$ for the normal distribution). In particular, the asymptotic correlation of extreme returns is equal to zero.

The correlation drops the more distant the returns are from the process mean. For extreme returns it goes to zero.

The question if correlation is higher in periods of volatile markets (large absolute returns) was usually examined by conditioning the estimated correlation on the observed (ex post) realizations of market returns. Dividing a bivariate sample in 50% large and 50% low returns[3] and assuming constant correlation between the series, the conditional correlation of large returns is certainly higher than the conditional correlation of low returns. This can be easily proved by simulation. However, Gibson, Boyer and Loretan (1999) show that conditional correlation is highly non-linear in the level of return the sample is conditioned.

Thus, the major source of differences comes from the way of conditioning. If conditioned on the absolute value of the realized returns, the conditional correlation increases with the threshold. This is because the truncated distribution retains the same mean as the total distribution. The estimated correlations are larger than the true correlation. However, if the sample is conditioned on signed extremes (positive or negative), which means that the mean of the signed distribution is not equal to the mean of the total distribution, then the conditional correlation of a multivariate normal distribution decreases with increasing threshold and finally reaches zero for extreme returns.

Therefore, it would be wrong to assume that extreme returns of dependent multivariate normal distributions appear highly correlated as they are large in

[2]The infinite lag polynomial was cut off at lag 500.

[3]According to their absolute size.

comparison with the mean of all returns. Two extreme returns are not necessarily correlated as they might not be large compared to the mean of the extreme returns.

Table 12.10 displays the correlations obtained for both the tails and the centers of the samples. It can be observed that correlations among center observations seem to be more balanced as the are more stable even beyond neighboring rating grades. Correlations in the tails are rather lower than correlations between the whole sample and positive almost only between neighboring series. Between neighboring investment grade series, tail correlations are higher and remain stable when the threshold increases, and they drop very late. To demonstrate this, a more detailed look is held at the tail correlations between neighboring rating grades of equal maturity.

For illustration, figure 12.4 provides scatterplots of the $u_{i,t}$-representations of the neighbored A and BBB for maturities 2 years, 10 years, and 30 years. The graphs below plots the corresponding tail correlations for increasing thresholds. While the tail correlations drop evenly for the 2-year and 30-year maturities, the behavior of the 10-year maturity tail correlation remains relatively high with increasing threshold first and suddenly plunges.

For the tail correlations, it can be concluded:

- If ever - tail correlations only play a role for neighboring rating grades.
- Between distant rating grades the tail correlations are sometimes not significant. For example, tail correlations between speculative and investment grade series are mostly weak and even become negative at lower thresholds. This is due to the fact that the corresponding tail observations of two series are often located on the opposite sides of their respective means (i.e. the means of the tail observations of the series).
- However, even neighboring rating grades show a rapid decline in tail correlations when the threshold increases.
- Only the $u_{i,t}$ of neighboring 10-year investment-grades[4] have tail correlations that are greater than their center correlations, and that drop relatively late.

Thus, it can be concluded that, at least for the most cases here, the correlations in the tails decline for increasing threshold. A systematic increase of correlations in the tails caused by outliers could not be observed. This is important as now the dependence can be modeled by methods based on the Gaussian correlation matrix without further adjustments of the time series. Such methods that describe the dependence between random variables with stable marginals based on the

[4] This means the rating grades AA - BBB.

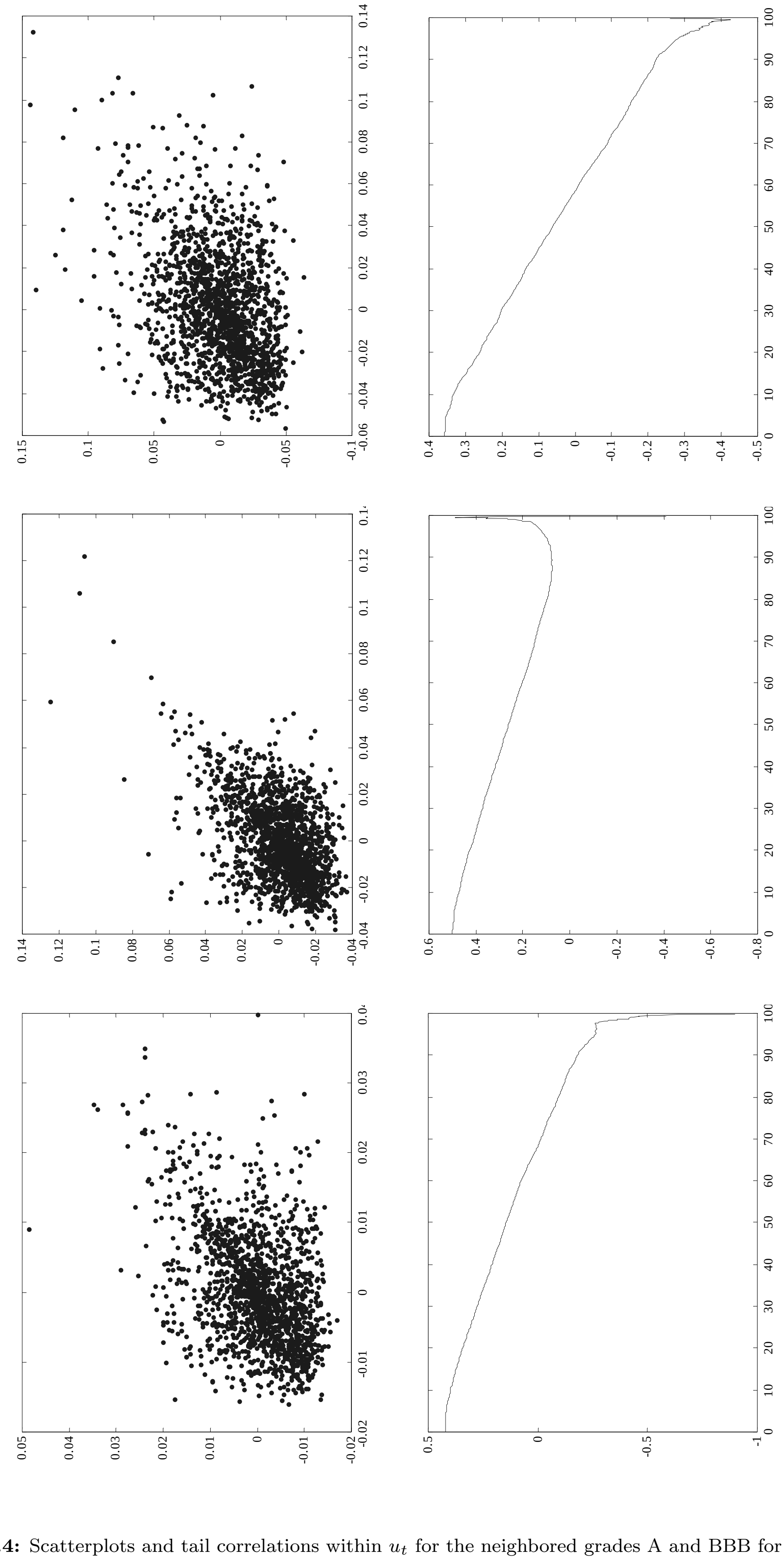

Figure 12.4: Scatterplots and tail correlations within u_t for the neighbored grades A and BBB for maturities 2 years (left), 10 years, and 30 years.

Gaussian correlation matrix are, for example, the above introduced sub-Gaussian vector and the *Gaussian copula with stable marginals* introduced in section 12.7.2. A truncation of the outliers in the tails of the sample is not required for the application of these methods.

With the model

$$(z_t - \mu_z)(1 - L)^d = u_t, \qquad (12.5)$$
$$\text{where } z_t = |\epsilon_t|^{1/2},$$

the LRD and the strong heavy-tailedness are captured, and thus do no more appear in the u_t.

The elements of u_t are i.i.d. and follow associated skewed stable distributions with common stability index, $S_\alpha(\beta, \gamma, \mu)$. Their stability index is mostly close to 2.

The dependence between the elements of u_t can, for example, be modeled by splitting each stable random variable $u_{t,i}$ into two independent components: a totally skewed stable and a symmetric stable random variable. The symmetric component models the dependence between the time series of different rating grades by the application of a sub-Gaussian vector. However, the application of such a sub-Gaussian vector is quite awkward for the given problem.

An advanced method of modeling the dependence between the elements of a stable random vector is the application of *copulas*. Copulas describe the dependence within a random vector much better than the correlation as a traditional measure. However, so far the use of copulas for the multi-asset case seemed to be too complex, especially for the stable non-Gaussian case. Examples have mostly just considered the bivariate or low-dimensional cases.

Considerable progress has been made in the development of copulas recently. The reader is here - for example - referred to Embrechts et al.[5] A big step forward in describing the dependent behavior of multi-asset returns under the stable non-Gaussian assumption has just been developed by Bravo Risk Management Group.[6]

12.7 Copulas As Measure For Dependence

The commonly used covariances are only one particular measure of stochastic dependence and are based on the assumption of multivariate normally distributed returns. The application of covariances respectively correlations becomes problematic in case of heavy-tailedness and skewness. Under the α-stable assumption,

[5] See Embrechts, McNeil and Straumann (1999) as well as Embrechts, Lindkog and McNeil (2001).

[6] Bravo-Group (2002).

for symmetric heavy-tailed returns sub-Gaussian vectors are able to model the dependence based on Gaussian covariances. Stable random vectors with asymmetric distributions so far could not be directly represented with their dependence structure. They had to be split into a symmetric dependent and a skewed independent component.

Currently copulas are seen as the best approach understanding the stochastic dependence between random variables.[7] The following section explains the general idea of copulas. Finally, a copula is introduced that describes the dependence for skewed stable non-Gaussian marginals.

12.7.1 General Idea of Copulas

This section explains the theory of copulas and provides the general formal concept.

Copulas deal with probability integrals and quantile transforms. Thus, before defining the concept of copulas, an important proposition will be illustrated first:[8]

Proposition 1 *Let X be a random variable with distribution function F. Let F^{-1} be the quantile function of F, i.e.*

$$F^{-1}(\alpha) = inf\,\{x|F(x) \geq \alpha\}\,,\ \alpha \in (0,1) \tag{12.6}$$

Then

1. *For any standard-uniformly distributed $U \sim U(0,1)$ there is $F^{-1}(U) \sim F$. This gives a simple method for simulating random variates with distribution function F.*

2. *If F is continuous then the random variable $F(X)$ is standard-uniformly distributed, i.e. $F(X) \sim U(0,1)$.*

The dependence between the real-valued elements of a k-dimensional random vector $(X_1, ..., X_k)$ is described by their joint distribution function $F(x_1, ..., x_k) = P\left[X_1 \leq x_1, ..., X_k \leq x_k\right]$. The random vector X could be transformed componentwise to have a standard uniform marginal distribution, $U(0,1)$. Assuming $X_1, ..., X_k$ have continuous marginal distributions $F_1, ..., F_k$ so that this can be achieved by a probability-integral transformation function:

$$T : R^k \rightarrow R^k, (x_1, ..., x_k)^t \rightarrow (F_1(x_1), ..., F_k(x_k))^t \tag{12.7}$$

[7] Embrechts, McNeil and Straumann (1999).

[8] Embrechts et al. (1999).

The joint distribution function C of $(F_1(X_1), ..., F_k(X_k))^t$ is then called the *copula* of the random vector $(X_1, ..., X_k)^t$ or the multivariate distribution F. It follows that

$$\begin{aligned} F(x_1, ..., x_k) &= P\left[F_1(X_1) \leq F_1(x_1), ..., F_k(X_k) \leq F_k(x_k)\right] \\ &= C(F_1(x_1), ..., F_k(x_k)). \end{aligned} \tag{12.8}$$

Definition. A copula is the distribution function of a random vector in R^k with uniform-$(0,1)$ marginals. Alternatively, a copula is any function $C : [0,1]^k \to [0,1]$ which has the three properties:

1. $C(x_1, ..., x_k)$ is increasing in each component x_i.
2. $C(1, ..., 1, x_i, 1, ..., 1) = x_i$ for all $i \in \{1, ..., k\}$, $x_i \in [0,1]$.
3. For all $(a_1, ..., a_k)$, $(b_1, ..., b_k) \in [0,1]^k$ with $a_i \leq b_i$ there is : $\sum_{i_1=1}^{2} ... \sum_{i_k=1}^{2} (-1)^{i_1+...+i_k} C(x_{1i_1} ... x_{ki_k}) \geq 0$, where $x_{j1} = a_j$ and $x_{j2} = b_j$ for all $j \in \{1, ..., k\}$.

Property 2 comes from the fact that the marginals are uniform-(0,1). Property 3 is true because the sum can be interpreted as $P\left[a_1 \leq X_1 \leq b_1, ..., a_k \leq X_k \leq b_k\right]$, which is non-negative.

Copulas are a natural way to measure the dependence between random variables. The properties of copulas are invariant under strictly increasing transformations of the underlying random variables. Linear correlation is not a copula-based measure and can often be quite misleading. It should not be taken as canonical dependence measure (Embrechts, McNeil and Straumann, 1999).

The concept of tail dependence relates to the amount of dependence in the lower left and upper right quadrant tail of a bivariate distribution. Tail dependence especially becomes important for extreme values. It has to be mentioned that tail dependence between two bivariate random variables is a copula property, and the amount of tail dependence is invariant under strictly increasing transformations of X and Y.[9]

12.7.2 Gaussian Copula with Stable Marginals

As the analysis of the correlations in section 12.6 evidence that significant tail correlations are mostly not seen, it makes sense to argue for the application of a *Gaussian copula* for the stable random vector u_t.

[9] Copulas such as the T-copula can model tail dependence but they are relatively complex. The Gaussian copula, however, is not suitable for tail dependence but is less complicated.

For their risk management software Cognity, Bravo Risk Management Group propose the application of a Gaussian copula with stable marginals. It is especially tractable for high-dimensional stable random vectors with even skewed stable marginals.

Unlike for the sub-Gaussian vector approach, the copula approach does not require the elements of the stable random vector to be split into a symmetric dependent and skewed independent component.

The procedure for the implementation of Gaussian copulas for stable random vectors is outlined in the following:[10]

Having an I-dimensional vector of random variables, X, where $X_{i,j}$ is the j-th observation of the i-th element of X, where $i = 1...I$ and $j = 1...J$. The X_i have stable-non-Gaussian marginal distributions. However, their dependence will be described by a Gaussian copula now. The copula is built as follows:

1. For each observation $j = 1...J$ of a random-vector element X_i, the value of the cumulative density function is estimated: $U_{i,j} = S_i(X_{i,j})$, $i = 1...I$, where S_i is the fitted stable cumulative distribution function (CDF) for the i-th random variable. $U_{i,j} = S_i(X_{i,j}) \in U(0,1)$.

2. For each set of observations j an I-dimensional multivariate vector N_j of a multivariate normal distribution is constructed, with components $N_{i,j}$. The i-th sample has a normal distribution $N(m_i, v_i)$ with mean m_i and variance v_i. m_i and v_i are the sample mean and the sample variance estimated from the sample $X_{i,j}$, $j = 1...J$. The I-dimensional multivariate vector N_j with Gaussian marginals is obtained by transforming the I-dimensional vectors U_j: $N_j = (\Phi_1^{-1}(U_{1,j}), ..., \Phi_I^{-1}(U_{I,j}))$, where Φ_i is a CDF of $N(m_i, v_i)$. Assuming that the real dependence is described by a Gaussian copula, this vector will be a multivariate normal vector.

3. With the N_j, $j = 1...J$, a multivariate normal distribution is fitted, $N(., \Sigma)$.

4. To simulate scenarios, samples are drawn from the multivariate $N(., \Sigma)$

5. The current Gaussian marginals have to be transformed into stable ones. Thus, each coordinate of each draw is converted to $U(0,1)$ by $W_j = (\Phi_1(N_{1,j}), ..., \Phi_I(N_{I,j}))$. The transformed simulations are denoted with $W_{i,j}$.

6. For each $i = 1..I$, $S_i^{-1}(W_{i,j})$ is constructed. The result are simulated multivariate random vectors with stable marginals and Gaussian copula.

[10] For a description in detail the reader is referred to the technical document of Cognity by Bravo-Group (2002).

The key advantage of this type of copula is the simplicity of its application, especially for high-dimensional random vectors.

For stable distributed asset returns, the T-copula has been used in so far in publications. The Gaussian copula could not be applied due to the tail correlations present between different assets. However, the T-copula that can model such tail correlations is quite complex and more difficult to implement than the Gaussian copula, which is much more comfortable to be used. The long-memory model of the developed credit return model and the resulting u_t allow to fit the Gaussian copula. This is a key finding of this chapter.

12.8 Value at Risk Estimation For The Long-Memory Model

Recalling the FARIMA(0,d,0) model for the $|\epsilon_t|^{1/2}$, where the innovations are described by the long-memory model:

$$(|\epsilon_t|^{1/2} - \mu_z)(1 - L)^d = u_t. \tag{12.9}$$

The way the Value at Risk caused by the innovations is determined differs from the common way as the model (12.9) forecasts the absolute values of the residuals, $|\epsilon_t|^{1/2}$. This is different from the way Value at Risk is determined for GARCH-type models as discussed in chapter 9.

In order to evaluate and compare the VaR accuracy of both EWMA and the long-memory model with an in-sample test for one-step-ahead VaR forecasts, it is useful to focus on the examination of isolated volatility models for the $\epsilon_{i,t}$ only - instead of the whole cointegrated VaR.

However, for the long-memory model the VaR is forecasted in a different manner. $|\epsilon_t|^{1/2}$ is forecasted by

$$|\epsilon_t|^{1/2} = L^d \left(|\epsilon_t|^{1/2} - \mu_z\right) + \mu_z + u_t. \tag{12.10}$$

The probability distribution of the forecast $|\epsilon_t|^{1/2}$ is determined by the probability distribution of u_t.

The estimation of the VaR for the fractional model and the counting of the exceptions are done in the following way: The forecasted variables are the elements of the vector of the absolute values $|\epsilon_t|$, however, not the signed elements $\epsilon_{i,t}$. The processes driving the $|\epsilon_{i,t}|$ are not i.i.d. because of the long-memory present in the absolute values. On the other hand, the stochastic process determining the sign of the $\epsilon_{i,t}$ is i.i.d. Thus, on each stage of the process, the probability for the sign in the following period is 0.5 for both "-" and "+".

While the distribution of $|\epsilon_t|^{1/2}$ is skewed to the right, the distribution of $sign(\epsilon_t)|\epsilon_t|^{1/2}$ is symmetric. The skewness is caused by the use of absolute values. For the variable $|\epsilon_t|^{1/2}$, both positive and negative extreme values of ϵ_t fall in the right tail of the distribution.

Studying the plot of one-step ahead forecasts of the $|\epsilon_t|^{1/2}$, it can be seen, despite of large residuals, that the path of the forecasted $|\epsilon_t|^{1/2}$ captures the movements of the process' volatility surprisingly well.

The probability forecast of the $|\epsilon_{i,t}|$ is driven by the probability distribution of the $u_{i,t}$. Assuming that the (conditional) probability of positive and negative values of $\epsilon_{i,t}$ is always 0.5, the VaR for the long-memory model is obtained as follows:

Assuming the probability that the absolute value $|\epsilon_{i,t}|$ exceeds a (positive) level C is α. Thus, with the above assumption that the probability of positive and negative signed returns is 0.5 each, the probability that a loss of C is exceeded in absolute value by a negative $\epsilon_{i,t}$ is $\alpha/2$.

Estimating, for example, the 90% quantile of $|\epsilon_{i,t}|$, it refers to the 95% quantile of the distribution of $\epsilon_{i,t}$. Estimating C as the $1-\alpha$ quantile of $|\epsilon_{i,t}|$, all negative observations whose absolute value is greater than C are counted as exceptions. The unconditional coverage is

$$\hat{\alpha}/2 = \frac{\#exceptions}{n} \ . \tag{12.11}$$

n is the observed sample size.

Aside from the given model $(|\epsilon_t|^{1/2} - \mu_z)(1-L)^d = u_t$, which is referred to as *model 1*, three other models for ϵ_t will be evaluated simultaneously:

- The $|\epsilon_t|^{1/2}$ are assumed to be i.i.d. and follow a stable law, referred to as *model 2*.
- The ϵ_t are assumed to be i.i.d and follow a stable law, referred to as *model 3*.
- Finally, the multivariate stable EWMA model of chapter 9, referred to as *EWMA model*.

The vector of the skewed u_t in *model 1* is fitted with the initial restriction $\beta = 1$. That means that even those vector elements that were found with $\alpha_i = 2$ originally can now be fitted by *skewed* stable distributions. This is important in order to capture the tail well.

The four elements 2y AA, 2y BBB, 10y A, 30y A of vector u_t originally exhibit a stability index α_i equal to 2 (Gaussian) if they are fitted without restrictions.

The obtained unconditional coverages in case of $\alpha_i = 2$ for the u_t series of 2y AA, 2y BBB, 10y A, 30y are displayed in parenthesis. It can bee seen that the empirical VaR is underestimated by the Gaussian distribution as it lacks both heavy-tailedness and skewness. The unconditional coverage clearly exceeds the theoretical value in this case.

The unconditional coverage for *model 1*, model 2, and *model 3* are given in tables 12.11 - 12.19. The results for the *EWMA model* are known from chapter 9.

The results of the long-memory model *model 1* are presented in tables 12.11 - 12.13.

For *model 2* (i.i.d. distributed $|\epsilon_t|^{1/2}$) and *model 3* (i.i.d. process with stable distributed signed residuals ϵ_t) the results of the unconditional coverage are given in table 12.14 and 12.19).

The results in tables 12.11 - 12.19 demonstrate that *model 1* outperforms *model 2* and *model 3* in terms of one-step-ahead VaR. More important, *model 1* also seems to be better than the *EWMA model* of chapter 9. This can be especially seen when calculating the *mean-absolute deviation* and the *mean-squared deviation* of unconditional coverage from the theoretical values over all rating / maturity combinations. This is shown in tables 12.20 and 12.21.

12.9 Conclusion

The VaR estimates of *model 1* over all 15 combinations of rating grade and maturity show on average the best results for unconditional coverage for the 95% and 99% VaR compared to the three other competing models.

Especially interesting is the comparison with the accuracy of the the stable EWMA which has been chosen as the favorite model among the traditional volatility models in chapter 9. The long-memory *model 1* clearly outperforms the stable EWMA model in terms of overall accuracy, as demonstrated by the mean absolute deviation and mean squared deviation in tables 12.20 and 12.21.

The application of the skewed stable distribution for the u_t performs extremely good for both the 95% VaR as well as for the higher quantiles 99% and 99.5% VaR.

The proceeding in chapter 12 can be summarized as follows:

- The FARIMA model describing the $|\epsilon_t|^{1/2}$ has been specified and the parameters have been fitted.
- With this model, the 95%, 99% and 99.5% one-step ahead Value at Risk are determined and its forecast accuracy has been compared with the forecast accuracy of three other models:
  - The $|\epsilon_t|^{1/2}$ are i.i.d. and follow a stable law, referred to as *model 2*.
  - The ϵ_t are i.i.d. and follow a stable law, referred to as *model 3*.
  - The stable EWMA model of chapter 9, referred to as *EWMA model*.
- Furthermore, a Gaussian copula with stable marginals is introduced to describe the dependence between the elements of the given stable random vector u_t.

The *key* findings derived in this chapter are:

- The best specification for the FARIMA model describing the $|\epsilon_t|^{1/2}$ is (0,d,0). As found by the moving-blocks bootstrapping test, the parameter estimates for the d_i are all significant. Except for the 10 year BB, they are even highly significant. This also confirms the choice of this model to describe the behavior of the credit returns. In addition, the robustness of the estimator developed in the former chapter 11 has been demonstrated.
- Analyzing the forecast accuracy for the one-step-ahead Value at Risk at the 95%, 99% and 99.5% level, the long-memory model clearly *outperforms* its competitors, among them the multivariate stable EWMA model as the preferred choice among the volatility models discussed in chapter 9.

- Modeling the dependence among the elements of a stable random vector, the application of a sub-Gaussian vector exhibits some practical problems: for example, the requirement of a common stability index for all vector elements, the splitting of each element into a dependent symmetric and an independent skewed component, or the estimation of the correlation based on truncated samples. *Copulas* are a better but often too complex way to model the dependence.[11] However, section 12.7.2 exhibits that the dependence between the elements of the stable random vector u_t can be fitted by a Gaussian copula as tail correlations are almost not present. Thus, a *Gaussian copula with stable marginals* is introduced which is less complicated and burdensome in the high-dimensional case than, for example, the T-copula.

The chosen multivariate stable FARIMA (0,d,0) demonstrates that a long-memory model can be parsimonious on the one hand while showing great flexibility on the other hand. Furthermore, despite of the stable assumption it has a tractable estimator (modified CSS). In terms of forecast accuracy it has outperformed the multivariate stable EWMA as the former benchmark.

Another advantage of the stable FARIMA(0,d,0) process is its self-similarity.[12] Self-similar processes can be easily scaled by the application of the *self-similarity property*. This property is beneficial when it comes to extend the forecast horizon of the long-memory model from one-day forecast periods to longer forecast periods.

Furthermore, the comparison of *model 1* with the two i.i.d. stable models *model 1* and *model 2* also demonstrates that the properties of the phenomenon *long memory* cannot just be explained by the heavy-tailedness of the process.

Chapter 13 provides an outlook on possible future applications of the credit return model developed in this thesis. It explains state-of-the-art methods for portfolio optimization based on Value at Risk measures and proposes a way to integrate the credit return model into a simulation-based portfolio optimization framework.

[11] See, for example, the T-copula.
[12] See definition in chapter 5.

2 year:					
Center:	AA	A	BBB	BB	B
AA	1	0.3136	0.3656	0.2102	0.0833
A	0.3136	1	0.3421	0.2275	0.0826
BBB	0.3656	0.3421	1	0.2253	0.0952
BB	0.2102	0.2275	0.2253	1.0000	0.1892
B	0.0833	0.0826	0.0952	0.1892	1
Tails:	AA	A	BBB	BB	B
AA	1	0.2997	0.2002	-0.0583	-0.0555
A	0.2997	1	0.2884	-0.0295	-0.0869
BBB	0.2002	0.2884	1	-0.0105	-0.0862
BB	-0.0583	-0.0295	-0.0105	1	0.1376
B	-0.0555	-0.0869	-0.0862	0.1376	1
10 year:					
Center:	AA	A	BBB	BB	B
AA	1	0.43	0.3878	0.1785	0.1422
A	0.43	1	0.3559	0.1697	0.1504
BBB	0.3878	0.3559	1	0.1864	0.1731
BB	0.1785	0.1697	0.1864	1	0.3652
B	0.1422	0.1504	0.1731	0.3652	1
Tails:	AA	A	BBB	BB	B
AA	1	0.5139	0.4619	0.0674	-0.0191
A	0.5139	1	0.4334	0.0203	-0.0619
BBB	0.4619	0.4334	1	0.0529	-0.0969
BB	0.0674	0.0203	0.0529	1	0.1285
B	-0.0191	-0.0619	-0.0969	0.1285	1
30 year:					
Center:	AA	A	BBB	BB	B
AA	1	0.2725	0.3075	0.1486	0.1327
A	0.2725	1	0.3097	0.149	0.1224
BBB	0.3075	0.3097	1	0.1687	0.1461
BB	0.1486	0.149	0.1687	1	0.3599
B	0.1327	0.1224	0.1461	0.3599	1
Tails:	AA	A	BBB	BB	B
AA	1	0.2296	0.2313	-0.1137	-0.0666
A	0.2296	1	0.218	-0.0711	-0.0105
BBB	0.2313	0.218	1	-0.1254	-0.1116
BB	-0.1137	-0.0711	-0.1254	1	0.0983
B	-0.0666	-0.0105	-0.1116	0.0983	1

Table 12.10: Center- and tail-correlations of the $u_{i,t}$.

	AA	A	BBB	BB	B
2y	0.0590 (0.0673)	0.0470	0.0554 (0.0644)	0.0512	0.0601
10y	0.0595	0.0524 (0.0578)	0.0584	0.0512	0.0530
30y	0.0560	0.0536 (0.0513)	0.0482	0.0500	0.0494

Table 12.11: Model 1: Unconditional coverage for 95% VaR. The results for the unrestricted fitting are in parenthesis.

	AA	A	BBB	BB	B
2y	0.0054 (0.0209)	0.0024	0.0024 (0.0238)	0.0060	0.0054
10y	0.0054	0.0060 (0.0131)	0.0060	0.0066	0.0054
30y	0.0018	0.0042 (0.0161)	0.0042	0.0054	0.0030

Table 12.12: Model 1: Unconditional coverage for 99% VaR. The results for the unrestricted fitting are in parenthesis.

	AA	A	BBB	BB	B
2y	0.0000 (0.0155)	0.0006	0.0000 (0.0137)	0.0012	0.0018
10y	0.0018	0.0024 (0.0083)	0.0018	0.0012	0.0000
30y	0.0006	0.0000 (0.0107)	0.0006	0.0006	0.0000

Table 12.13: Model 1: Unconditional coverage for 99.5% VaR. The results for the unrestricted fitting are in parenthesis.

	AA	A	BBB	BB	B
2y	0.0637	0.0488	0.0572	0.0458	0.0530
10y	0.0578	0.0506	0.0584	0.0476	0.0476
30y	0.0560	0.0566	0.0524	0.0458	0.0429

Table 12.14: Model 2: Unconditional coverage for 95% VaR.

	AA	A	BBB	BB	B
2y	0.0041	0.0023	0.0023	0.0011	0.0023
10y	0.0053	0.0053	0.0059	0.0023	0.0011
30y	0.0023	0.0053	0.0047	0.0005	0

Table 12.15: Model 2: Unconditional coverage for 99% VaR.

	AA	A	BBB	BB	B
2y	0	0.0006	0	0	0
10y	0.0017	0.0011	0.0029	0.0006	0
30y	0.0006	0	0.0006	0	0

Table 12.16: Model 2: Unconditional coverage for 99.5% VaR.

	AA	A	BBB	BB	B
2y	0.0513	0.0483	0.0453	0.0304	0.0369
10y	0.0572	0.0572	0.0447	0.0328	0.0340
30y	0.0548	0.0524	0.0507	0.0250	0.0268

Table 12.17: Model 3: Unconditional coverage for 95% VaR.

	AA	A	BBB	BB	B
2y	0.0036	0.0060	0.0030	0.0012	0.0036
10y	0.0030	0.0054	0.0066	0.0030	0.0036
30y	0.0024	0.0060	0.0024	0.0012	0.0018

Table 12.18: Model 3: Unconditional exceptions for 99% VaR.

	AA	A	BBB	BB	B
2y	0.0012	0.0012	0.0012	0.0012	0.0006
10y	0.0012	0.0006	0.0006	0.0018	0.0012
30y	0.0000	0.0012	0.0012	0.0000	0.0012

Table 12.19: Model 3: Unconditional exceptions for 99.5% VaR.

	model 1	model 2	model 3	Stable EWMA
95% VaR	0.0044	0.0051	0.0100	0.0102
99% VaR	0.0054	0.0069	0.0065	0.0061

Table 12.20: Mean absolute deviation from the theoretical value of unconditional coverage, for both 95% VaR and 99% VaR.

	model 1	model 2	model 3	Stable EWMA
95% VaR	3.0145	3.7617	16.391	18.122
99% VaR	3.0960	5.2740	4.4816	4.1133

Table 12.21: Mean squared deviation from the theoretical value of unconditional coverage, for both 95% VaR and 99% VaR. The values have to be multiplied with e^{-5}.

Chapter 13

Outlook - Further Applications Of The Credit Return Model

This chapter describes further applications where the credit return model should proove useful:

- determination of the VaR for a portfolio of individual corporate bonds.
- optimization of the portfolio risk by restructuring its positions.

In detail, the chapter covers three major topics:

- It sketches the modeling of the price path of *individual* corporate bonds in relation to the credit return model.
- It proposes a simulation framework to obtain the VaR of a corporate bond portfolio.
- It discusses the state of the art on VaR optimization by restructuring the positions of a given portfolio with respect to the application of stable distributions for asset returns.

13.1 Proposal For Bond-Portfolio VaR

This section sets out the application of the cointegrated VAR model developed in the previous chapters to describe the returns of *individual* corporate bonds.

A possible solution of linking the risk of an individual corporate bond to the cointegrated VAR credit return model is given. By simulating future credit returns for a given time horizon, the VaR of the portfolio of corporate bonds is computed.

However, measuring the VaR of a corporate bonds' portfolio is only one step in the risk management process. The simulation framework developed in this thesis may also form a basis for *scenario analysis* or *portfolio optimization* as well. The latter becomes an issue once investors are no longer satisfied with the risk of their portfolios.

Depending on their risk appetite and the funds available for covering economic capital, the investors might prefer to restructure the portfolio - for example in order to obtain a lower VaR or a different risk-return profile. Thus, this chapter also outlines a brief overview on portfolio optimization based on VaR measures (the discussion of optimization procedures is thought as an outlook and not considered to be a substantial part of this work).

13.1.1 Simulation-Based Risk Measurement

Simulation-based tools are commonly applied for scenario analysis of a given portfolio. In particular, they provide additional insights:

- When the portfolio contains non-linearities.
- When market distributions are not normal.
- Or when there are multiple horizons.[1]

Complex models such as the cointegrated VAR for credit returns developed in this thesis require computational simulation methods to determine the VaR for a future time horizon which takes the impact and interdependence of all risk factors into account.

In order to determine the VaR of a given portfolio, a set of scenarios has to be simulated with the credit return model for the desired time horizon. For each scenario, the value of all single assets, and thus the value of the whole portfolio, is then calculated. With the portfolio value for each scenario, the empirical distribution of the future portfolio value is obtained. The VaR is taken as a defined quantile of this distribution.

13.1.2 Modeling The Risk Of A Single Bond-Position

The credit return model describes the path of the average log prices for a given credit grade. The average credit risk within a given rating grade is assumed to remain relatively stable in the long term because it is derived exclusively from corporate bonds of very similar credit quality. However, the average price difference between the rating grade and the treasury bond may change due to a number of

[1] See Mausser and Rosen (1999).

effects. During an economic recession, for example, the market assumes a weaker credit quality for a certain number of bonds within a rating grade, and the average log price of the rating grade thus declines while the rating agencies do not change the credit rating of the firms affected. The underlying idea is that economic cycles affect the average probabilities of default within a rating grade. Certainly, such effects are limited, as bonds whose spread changes dramatically compared to the other bonds of the same rating grade will sooner or later be up- or downgraded by the rating agencies, then having no further influence on the average spread of their former rating grade.

Market trends and liquidity effects in the corporate bond market typically have an impact on each bond of a given rating grade, changing the average credit spread of that grade. For example, if investors prefer to switch to treasury quality, then liquidity is withdrawn from the corporate bond markets. The spreads of the corporate bonds therefore increase.

Rating agencies provide annual transition matrices giving the probabilities of firms migrating from one rating grade into another. However, such transition matrices neither contain information about when the switch in rating happens nor reveals the price path a corporate bond follows during a rating change. So, when the rating change is announced, the spread of the bond might not change because the market has already anticipated the change in credit quality. Thus, when dealing with forecasts of VaR for short horizons of 10, 20, or 60 days, such transition matrices are not very helpful in drawing conclusions to the potential spread or price changes. Instead, it is essential to focus on the price processes of the corporate and treasury bonds.

The approach presented here identifies three main drivers for the price changes of a corporate bond:

- The movements of the treasury yield curve.
- The movements of the average spread of the given rating grade.
- Individual effects which might be mainly due to changes in the credit quality of the issuer.

Thus, the next step is to develop a mechanism to describe the price movement of an individual corporate bond on the basis of the cointegrated VAR credit return model.

A credit analysis based purely on rating migration and default events would not allow a valuation on a daily basis. The price path a certain corporate bond takes when its credit rating changes hinges on the individual circumstances and information available to the market before and during the rating change. In many cases the market anticipates changing credit quality before the rating agencies react.[2]

Therefore, for individual corporate bonds, annual rating migration probabilities are not a good indicator for the short-term price behavior.

How can up- and downside risks of individual bonds be captured? From this, again the question arises how a link between the credit return model and the behavior of individual corporate bonds may be built.

First, the behavior of individual corporate bonds with given rating has to be analyzed and compared to the average of the corresponding rating grade - assuming that the "average bond" has exactly the same specifications (coupons, coupon dates, maturity) as the individual bond.

Next follows the construction of the individual bond's price under the assumption that its yield follows the average of its rating grade. This is done by taking historical price series of the rating grade's averages for 1 month, 2 years, 10 years, and 30 years maturity.

For the series in the credit return model, the time to maturity for each observation day is the same (constant time to maturity with 1 month, 2 years, 10 years, 30 years), whereas the individual corporate bond's time to maturity declines from day to day.

In order to compute the historical daily log prices of a corporate bond which obeys exactly the yield curve of its rating grade average, the following procedure is applied:

- The time from t to each component C_i of the bond's cash flow has to be measured for each day t. This time is denoted by T_i.
- For each component C_i, its present value PV_i at t is determined by interpolation between the log prices of the known maturity points 1 month, 2 years, 10 years, and 30 years at t. Here, a cubic spline interpolation is applied.[3]
- The present value of the corporate bond - as if it were following the average yield to maturity of its rating grade exactly - is obtained by summing the PV_i.

Plotting both price paths, it is often the case that the observed price of a bond does not match the average price of its rating grade.

Figure 13.1 illustrates that the log price of an individual corporate bond and the log price of its rating grade average often move in parallel. Thus, the individual corporate bond's price path usually lies between the averages of two neighboring rating grades.

[2]See also Beck (2001). Research on the problem of rating agency announcements was also done by citep*HaHL92.

[3]Burden and Faires (1997).

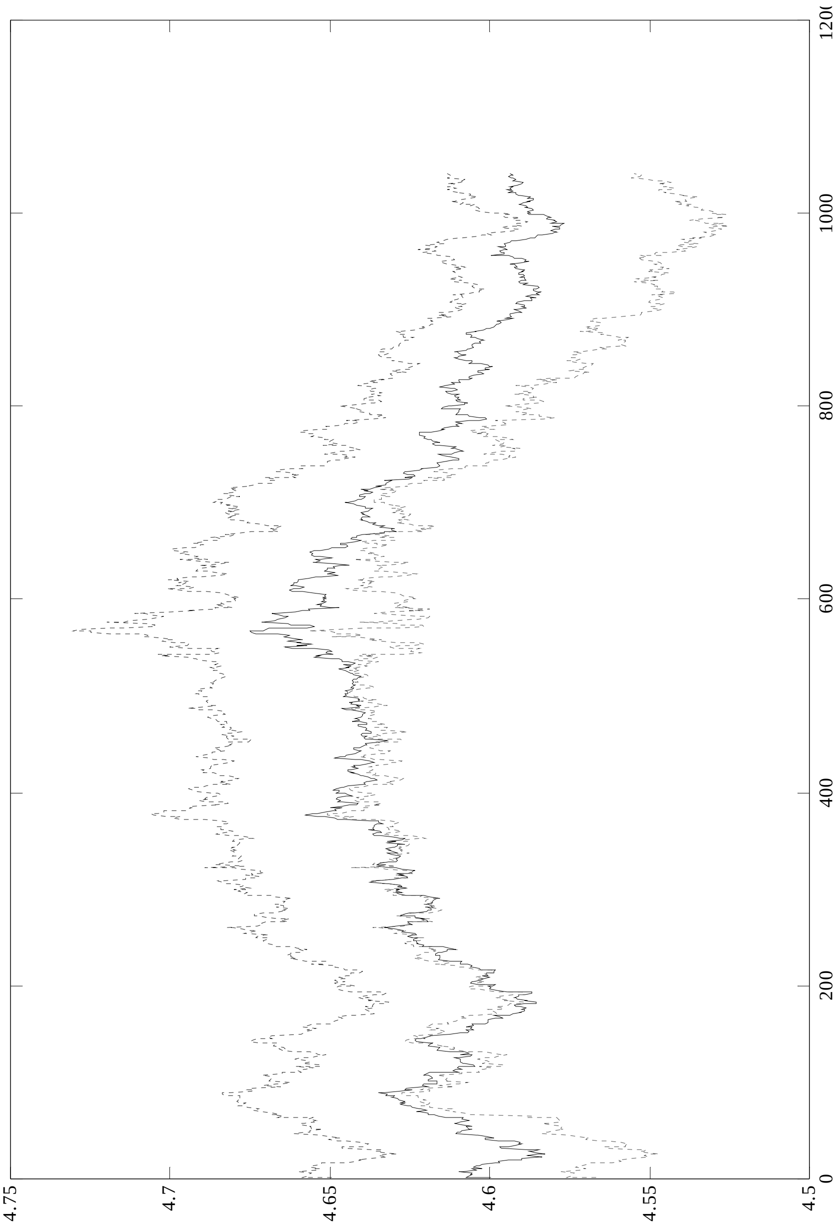

Figure 13.1: Plot of log prices of an individual bond and neighboring average rating grades (dotted lines).

The idea is to express the movement of the individual bond's log price $log(p_{IndB})$, which can be expressed by the movement of the five rating grades of the credit grade model with the same maturity as the individual bond.

$$log(p_{IndB,t}) = g_1(t)log(p_{1,t}) + g_2(t)log(p_{2,t}) + g_3(t)log(p_{3,t}) + g_4(t)log(p_{4,t}) \\ +g_5(t)log(p_{5,t}) + g_6(t)log(p_{6,t}). \quad (13.1)$$

. The $g_i(t)$, $i = 1...5$, are autoregressive processes of order 1 and sum up to a constant: $\sum_{i=1}^{5} g_i(t) = constant$.

Now those two of the $g_i(t)$ are selected, $g_{i1}(t)$ and $g_{i2}(t)$, which are most significant according to the Akaike Criterion. The selected i are i_1 and i_2.

The autoregressive process is then described by

$$g_{i1}(t) = constant - g_{i2}(t) \text{ and} \\ g_{i2}(t) = c_1 + c_2 g_{i2}(t-1) + a_t \ , \quad (13.2)$$

where a_t is a stable distributed random variable. $g_{i1}(t)$ and $g_{i2}(t)$ control the change in credit quality of an individual bond relative to the paths of the credit rating grade averages.

Such a model is designed to perform simulations of future returns for short periods of time (days). The fitting of (13.2) can be derived from the historical path of the individual bond.

For many corporate bonds available in the market, historical prices are often not available. They are often traded over the counter (OTC), and prices are not accessible because they are not published by the financial institutions involved. Other bonds are not traded every day, meaning that only a fragmentary price history is available.[4]

For the pricing of corporate bonds without price history, it is reasonable to assume for the given model that their prices exactly follow the average of their rating grade. Thus, their prices can be constructed.

13.1.3 Measuring Bond Portfolio Risk

After the $g_i(t)$ coefficients have been fitted, future scenarios can be simulated with the credit return models for 2 years, 10 years, and 30 years. The log prices for the one-month maturity are kept fixed for simplicity as they hardly move at all compared to the other maturities. With the credit return model, a specified corporate bond can now be valued as if it were following the movements of rating grades i_1 and i_2. There are three risk factors of the credit return model:

[4] Sometimes pricing sources exist, but their subscription is fairly expensive.

- The returns of the treasury bonds for each maturity.
- The common credit risk factor for each maturity.
- The vector of disturbance terms of the model that describes the innovations $\epsilon_{,t}$ for each maturity.

The credit return model simulates future representations for the combinations of maturities and credit ratings displayed in table 13.1:

Maturity/Rating	AAA	AA	A	BBB	BB	B
2 year	•	•	•	•	•	•
10 year	•	•	•	•	•	•
30 year	•	•	•	•	•	•

Table 13.1: Matrix of maturity and rating combinations.

For maturities lying between two maturity points of the model, a cubic spline interpolation is applied to obtain the corresponding return.

The introduced framework of the credit return model to obtain the VaR of a corporate bond portfolio is simulation based, making an optimization of the portfolio with to VaR as risk measure computationally very burdensome, especially for portfolios with a large number of positions.

The fact that three risk factors and the long-term relationship of the price paths are identified is a key advantage of this credit return model.

Due to the number of risk drivers, the credit return model makes it possible to extract a lot of information. This is helpful when performing *scenario analysis* and *stress testing*, for example.

This credit return model thus allows the returns and prices of *individual* bonds to be described. The link to the price process of the individual bond can be modeled by the relation defined in equations (13.1) and (13.2).

The topic of the following section is portfolio optimization. However, for portfolio optimization based on VaR as risk measure, a solution can only be obtained with this model by simulating scenarios. This is because an optimization with VaR causes difficulties due to some unpleasant properties of VaR on the one hand and the complexity of the model on the other.

The following section provides an overview of current state of the art in portfolio optimization and also covers the case when the returns follow a stable law. The section discusses alternative VaR measures and their applicability for optimization algorithms. While it is supposed to provide an overview on portfolio optimization, it is restricted in that it does not discuss a possible algorithmic solution in terms of the given credit return model.

13.2 Scenario Optimization - An Overview

The identification and measurement of the current portfolio risk are the initial steps in the risk management process. However, the investor or risk manager might not be happy with the current risk level. For example, the economic capital required to cover the potential losses might not be sufficient.

In order to compensate possible downside movements within a given horizon, the investor has to provide economic capital - usually measured by VaR. And since economic capital is limited, the risk manager is concerned with reducing the portfolio's VaR.

Alternatively, the investor might not be satisfied with the given risk-return ratio and desires to move the portfolio closer to the efficient frontier.[5]

In the following there is a brief sketch of the risk management process and scenario optimization under the VaR criterion. The objective is either (i) to minimize the portfolio's VaR or (ii) to optimize the ratio of mean return and VaR. An optimization algorithm simultaneously adjusts all the positions of the portfolio subject to defined restrictions.

13.2.1 The Risk Optimization Process

Risk management targets the identification, measurement, and the reduction of the sources of risk. The final goal is to obtain the desired risk profile for an investor. The process of credit risk optimization within a financial firm comprises steps on three levels:

- On the strategic level, the company has to optimize the weights of different credit product sectors.
- On the tactical level, the company has to decide about the weights for each obligor or classes of obligors.
- And third, the company optimizes the weights for each asset in a set of financial products. This happens on the operational level.

The complete process is illustrated in figure 13.2.

Risk optimization in order to obtain the desired risk profile for a specific investor is generally based on the portfolio theory developed by Markowitz (1952) and Sharpe (1964). It was originally applied to market risk with symmetric distributions of the returns. For credit instruments with possible defaults and rating migrations, the distributions of credit returns become skewed with fat tails. However, most of the time neither a default nor a severe change in credit quality can be observed. Thus, there is only very little observation of extreme downside events.

[5] See Markowitz (1952) and Sharpe (1964).

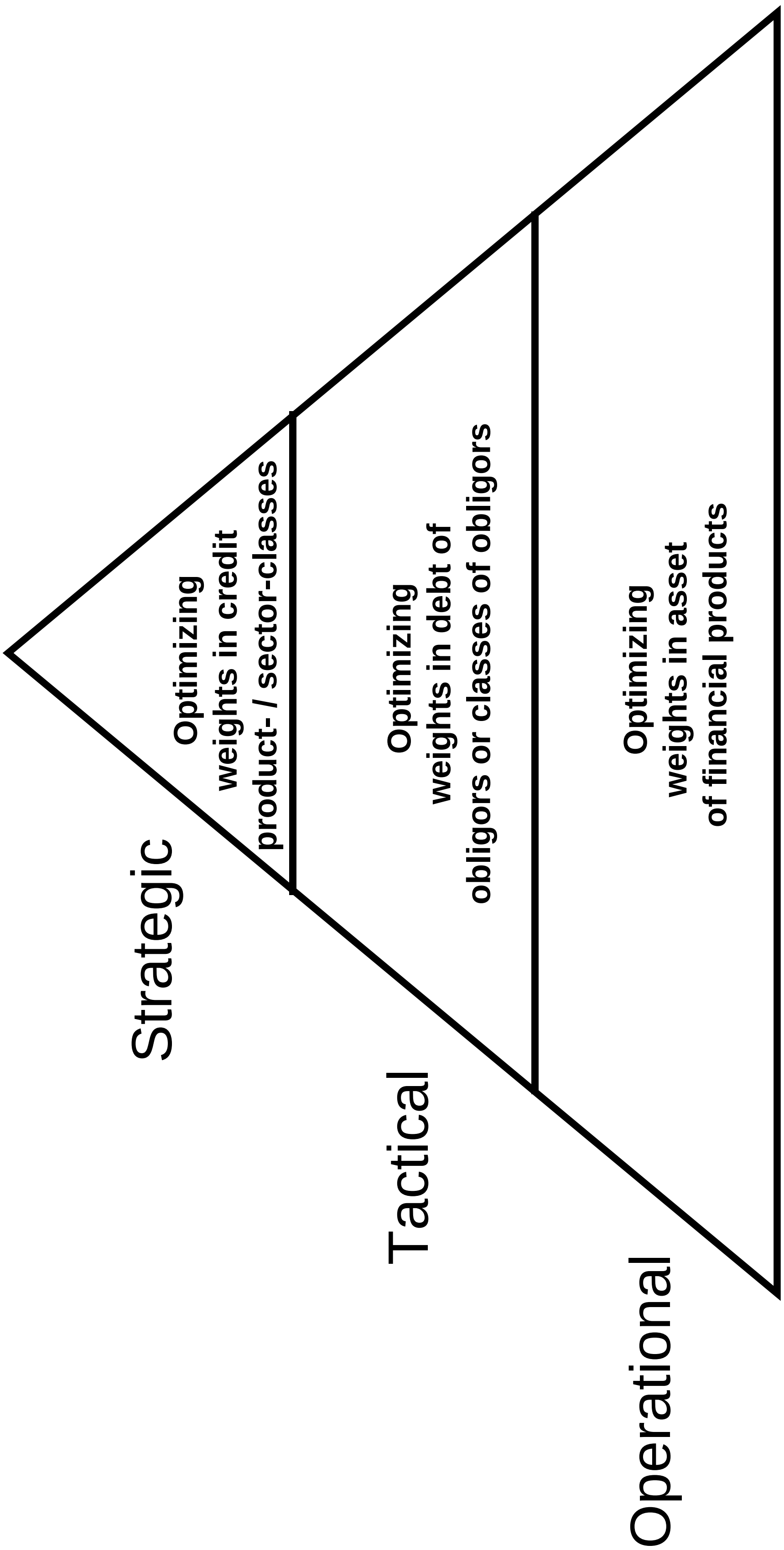

Figure 13.2: The three steps of credit risk optimization.

13.2.2 VaR Measures For Portfolio Optimization

Contrary to well-known mean-variance approaches, VaR-based portfolio optimization has recently experiened considerable development. VaR improves previously used risk measures for portfolio optimization, such as standard deviation. Standard deviation requires the Gaussian assumption for the returns; yet, the reality consists of fat-tailed and skewed distributions. On the other hand, standard deviation is relatively comfortable in terms of computational issues. VaR, however, requires intelligent computational methods in optimization.[6] Some approaches use mean-VaR and mean-variance portfolio optimization methods simultaneously.[7] However, VaR has awkward properties for optimization algorithms. Alternative risk measures to VaR which are applied for portfolio optimization are, for example, the *semi-standard deviation* and *conditional VaR*. Others are *expected regret* and *maximum regret*. It has been discovered that the resulting efficient frontiers are quite different depending on the chosen risk measure.[8]

The portfolio VaR is derived from simulated scenarios. Due to the large number of simulated scenarios, the size of the optimization problem becomes large as well.

Linear programming models have usually been applied to optimize portfolio risk.[9] But actually, there are very few papers dealing with the optimization of credit risk in portfolios.

However, there are some pitfalls in the choice of the quantile-based measure VaR as the variable to be minimized: conceptually, VaR could be handled as the optimizing measure of course, however, in practice it blows up the problem and makes it uncomfortable. In order to apply elegant linear programming and avoid the use of integer programming, VaR is not tractable. Arvanitis, Browne, Gregory and Martin (1998), for example, has applied a brute-force random search method. More advanced numerical methods can be applied with alternative risk measures that are more computationally efficient: For example, expected regret[10], maximum regret[11], or conditional VaR[12].[13] Another more tractable measure for portfolio optimization would be the so-called *nVaR*[14], which is also linear with

[6]Albanese, Jackson and Wiberg (2001).

[7]Wang (2000).

[8]Gaivoronsky and Pflug (2001).

[9]Mausser and Rosen (1999).

[10]Regret is defined as the difference between a scenario outcome and a benchmark. Expected regret is the expectation of losses that exceed a fixed threshold K.

[11]Maximum regret is the largest loss in excess of a threshold K

[12]Conditional VaR is the expected loss on a portfolio given the return is less than VaR.

[13]Mausser and Rosen (1998).

[14]The nVaR of a position is defined as the loss in a threshold scenario and is linear with the size of the position. The portfolio is optimized via the derivative of nVaR with respect to the given positions.

the number of units of each position in the portfolio and therefore reduces the computational effort.[15]

These alternative measures can be employed to reshape the distribution of credit returns, thus also improving quantile-based measures.

As mentioned, VaR has some undesirable properties: lack of sub-additivity, non-convexity, and non-smoothness. This was pointed out by Artzner.[16] For this reason, he suggested conditional VaR (CVaR) as an alternative measure.

CVaR has the properties subadditivity and convexity. When the return-loss distribution is normal, both CVaR or VaR result in the same optimal portfolio. However, credit returns are not normally distributed.

Stable non-Gaussian approaches. Tokat, Rachev and Schwartz (2001) introduce a solution methodology for the optimization of a portfolio of equity investments. The investor may choose between equity and cash. The model considers time variation in the expected returns on equity. The risk factors follow a stable non-Gaussian law. The solution is a multistage stochastic asset allocation problem with decision rules.

The authors have generated economic scenarios both with stable and Gaussian innovations. As reward measure they use the mean compound portfolio return. As risk measure they apply CVaR as an alternative measure of loss. Due to the lack of sub-additivity, the VaR of a portfolio with two bonds may be greater than the sum of individual VaRs. CVaR, however, has the sub-additivity property.

In their analysis they find that stable scenario modeling leads to asset allocations which are different by 20% compared to the normal assumption. This is due to the fact that the normal scenarios underestimate the risk. However, they note that the effect of fat-tailed returns on the asset allocation is contingent on the utility function and risk aversion level of the investor.

Ortobelli, Huber and Höchstötter (2001) also analyze the application of the stable distribution in portfolio choice theory. They examined empirical differences among the optimal allocations obtained with Gaussian and stable non-Gaussian distribution. They apply mean-risk analysis similar to Markowitz-Tobin's mean. As the stable distribution is more adherent to reality, stable methods improve performance measures. Ortobelli, Huber, and Schwartz also find significant differences in asset allocation between the sub-Gaussian and mean-variance model. The same effect was discovered by the authors when the investor uses VaR as measure for the risk. The impact of fat-tailed returns is even greater for these risk measures since they concentrate on the tail of the distribution.

[15]Mausser and Rosen (1998).

[16]See Artzner, Delbaen, Eber and Heath (1998) and Artzner, Delbaen, Eber and Heath (1999).

13.3 Summary

The first part of this chapter has proposed a model to describe the behavior of an individual corporate bond in relation to the credit return model.

The second part is a brief literature review about portfolio optimization under VaR and alternative risk measures for optimization algorithms. In addition, it summarizes two publications which perform asset allocation under the stable assumption.

For smaller tasks with a low number of positions in the portfolio, simulation-based optimization with VaR as risk measure might be tractable.

For a larger number of positions, however, alternative risk measures have to be chosen. Such measures have desirable properties for elegant optimization algorithms. Improving these measures generally leads to a corresponding improvement in VaR. Under the Gaussian assumption, both VaR and Conditional VaR obtain the same optimization results.

However, credit returns are better captured by stable distributions. The application of the stable distribution yields to different results for optimal asset allocation than obtained under the Gaussian assumption.

Chapter 14

Conclusion

14.1 Brief Summary

In this thesis, several phenomena present in time series of credit returns have been analyzed. On the basis of the results obtained, an appropriate model for the description of credit returns and the forecasting of VaR have been built.

Four phenomena of credit returns discussed are:

- Heavy-tailedness and peakedness.
- Time-varying volatility (volatility clustering).
- Long memory.
- Cointegration of different rating grades.

In chapter 3, the properties heavy tailedness, peakedness, and skewness have been examined for credit returns. It has been demonstrated that the modeling of the risk-factors with stable distributions outperforms the Gaussian models in terms of VaR. A modification of the credit model of Rachev, Schwartz and Khindanova (2001) has been developed in order to make its implementation and usage easier.

Stochastic processes that have the capability to describe financial prices as stable ARMA and GARCH processes have been introduced in chapter 4. In chapter 5, the theory of long-range dependence has been addressed. In addition, long-memory effects have been analyzed for returns of marketable credit instruments, under both the stable and Gaussian assumptions.

In chapter 6, the phenomenon that log prices of different rating grades show a behavior of cointegration is looked at. Chapter 8 proposes a model that captures this behavior: the stable cointegrated VAR. It has been applied to construct a model for credit returns that has the capability to describe the behavior of bond returns depending on their rating grade and maturity. To model the price chances for several rating grades simultaneously, a cointegrated vector-autoregressive approach was chosen since the price paths show similar trends in the long term. Three different maturities are selected, and a vector-autoregressive model is fitted for each maturity. The risk factors and innovations follow a stable law with $\alpha < 2$. The innovations of the model exhibit volatility clusters. In order to account for the clustering volatility in the returns, both stable GARCH(1,1) and stable EWMA are embedded and compared. The EWMA turned out to be more practical in its application and equally performant in terms of forecasting accuracy for VaR. As another result, both stable volatility models, GARCH(1,1) and EWMA outperform their Gaussian counterparts in VaR accuracy. The stable models are more conservative for the 99% VaR in particular.

Chapter 10 examines the innovation of the credit return model for LRD. One of the major findings of this chapter is that the presence of long memory is by far stronger and more significant in the absolute values of credit returns and not in the signed credit returns themselves. Based on this finding, the credit return model has been modified and a multivariate long-memory component has been developed. The long-memory model outperforms not only the multivariate stable EWMA model preferred in chapter 9 but also other competitors in terms of forecast accuracy for VaR.

The cointegrated vector-autoregressive model may be applied to simulate future scenarios of average bond prices subject to rating grade and maturity. The individual corporate bonds are priced based on the future scenarios generated. Chapter 13 sketches a proposal for modeling this. For individual bonds, future price paths can be simulated for the following cases:

- Corporate bonds with a rating grade (e.g. from Standard & Poors) that lacked a price history in the past.
- Corporate bonds with a price history available up to the present.

Furthermore, chapter 13 gives a brief overview on the current state of the art of scenario optimization, under both VaR and alternative measures.

14.2 Review Of Results

Chapter 1 explains the nature of credit risk and gives a brief sketch of the general types of credit risk models. A major handicap of common credit risk models is the Gaussian assumption. Thus, an alternative distributional assumption could help to better capture the heavy tails of credit returns. In the following the class of stable distributions is introduced. Furthermore, the credit risk model by Rachev, Khindanova, and Schwartz (2001) based on the stable assumption is reviewed. In this thesis a modified approach is developed by changing the definition of the credit returns. The modified model shows excellent results in forecasting VaR of corporate bond portfolios, and it can be implemented with less effort.

The implementation of the modified model requires solely historical prices of individual corporate bonds and historical treasury yield curve data but avoids the generation of historical yield curves for particular corporate bond rating grades.

The predominant performance of the stable model for VaR measures is compared to the application of the Gaussian distribution. The stable VaR performs excellently for the 99% confidence level while the Gaussian VaR largely underestimates the empirical VaR. The stable 99% VaR slightly overestimates the empirical 99% VaR. This usually leads to better VaR results as the empirical VaR often underestimates the true VaR due to a low number of samples in the tails.

To sum up, the heavy-tailedness and skewness property of credit returns are captured extremely well by the application of the stable non-Gaussian distribution.

AR(I)MA and GARCH processes have been discussed in chapter 4. These are commonly known processes that can be applied to describe financial returns. Furthermore, they are representatives of short-memory processes. Both AR(I)MA and GARCH processes are discussed under both the Gaussian and stable assumptions. GARCH processes explain clustering volatility. The volatility follows an autoregressive process and reacts to shocks of previous periods. This phenomenon finds its application in chapter 9.

Aside from heavy-tailedness and skewness, long memory is another phenomenon that has been observed for financial time series (e.g. stock returns) in the past. The thesis discusses this property in conjunction with credit instruments. In addition to the common distinction between I(0) and I(1) processes, fractional processes allow non-integer orders of integration. Fractional processes can be persistent or antipersistent. Persistent fractional processes are long-range dependent. A sign of long memory is the so-called burstiness of the plotted time series. Long-range dependence (LRD) is characterized by hyperbolically decaying autocorrelations, and the property that large/small representations are more likely to be followed by large/small representations. The chapter finishes with an empirical study of the presence of long memory in the returns of bond indices. Applying the methods Aggregated Variance and Absolute Values of Aggregated Series, all series exhibit a Hurst Exponent H greater than 0.5 which means LRD under the Gaussian assumption ($\alpha = 2$).

While the hypothesis *LRD is present* can be confirmed for two series at the 95% confidence level under the Gaussian assumption, under the stable assumption it can be confirmed for somewhat lower confidence levels. Thus, there is significance fin the presence of LRD in credit returns, although it is weak in its degree.

Starting with chapter 6, the thesis develops a framework for simulating future credit returns on the basis of credit rating and maturity. The framework is realized with a cointegrated vector-autoregressive (VAR) model which assumes that the variables follow a stable law. For each maturity, such a six-dimensional cointegrated vector-autoregressive model is built. Aside from the treasury returns there is another common risk factor: the residuals of the cointegrated VAR follow the residuals of the AAA equation. Thus, the residuals of AA, A, BBB, BB, and B have been regressed over the AAA residuals. As the remaining innovations exhibit time-varying volatility, they are modeled by a stable exponentially weighted moving average model, which has been found preferable to the GARCH(1,1) model.

The risk factors are the treasury returns with equal maturity, the *common* credit risk factor, and the individual innovations of each rating grade. The risk factors themselves are dependent over the ratings and the maturities. The treasury log prices are found as a mean-reverting process over time. For the cointegration vectors, the traditional relations found with the Johansen procedure proved to be too weak to prevent the log price paths of neighboring rating grades from intersecting. Thus, a more restrictive cointegration vector was applied.

The empirical results for one-step-ahead forecasts of VaR demonstrate that the stable GARCH(1,1) is outperformed by the stable EWMA, although the multivariate GARCH has many more parameters. In addition, the EWMA shows slightly better performance under MSE and MAE and almost identical performance in log likelihood. The advantage of EWMA is its simplicity since only one parameter has to be estimated for the model. Its tractability remains the same even when faced with an increasing number of equations while the number of parameters for multivariate GARCH(1,1) explodes, even for the constant correlation version. Therefore, for high-dimensional multivariate applications, EWMA is preferable.

Long-memory models extend the knife-edge choice of I(0) and I(1) processes to processes with non-integer order of integration. Unlike traditional volatility models which have been integrated into the vector-autoregressive model, long-memory models exhibit better forecasting properties for longer horizons. In addition, the calibration of traditional volatility models has to be updated frequently Mikosch and Starica (2000b) while such structural breaks over longer time spans are better captured by long-memory models. As mentioned, the LRD is found to be strong in the absolute values of the $\epsilon_{i,t}$. Thus, LRD is much stronger in absolute returns, which represent the behavior of volatility. This is demonstrated in section 10.3 with the MRS statistic. The multivariate stable FARIMA(0,d,0) proved itself to be the most suitable specification. Known alternative models for LM volatility have been found to be too complex for the high-dimensional multivariate case.

Comparing the forecasting accuracy of the long-memory volatility model with the forecasting accuracy of the traditional stable EWMA volatility model, it can clearly be seen that the long-range dependence (LRD) model shows better performance.

Tests evidence that the long-memory model forecasts VaR with greater precision than the competing volatility models, among them the stable EWMA model.

The advantages of the stable long-memory model can be summarized as follows:

- Within the cointegrated VAR, the long-memory model shows better accuracy in VaR forecasting over all ratings and maturities than its competitors. The estimates are conservative but do not exhibit too strong overestimation.
- The long-memory credit return model yields balanced performance for all rating grade / maturity combinations as it allows an individual fractional differencing parameter for each equation. However, it still remains parsimonious and of good tractability.
- The long-memory credit return model maintains the non-fractional cointegration relationships for the cointegrated VAR in order to limit the complexity of the multivariate model. LRD is isolated in the innovations.
- With the FARIMA(0,d,0) model, the extension to longer forecasting horizons can easily be obtained by scaling. The basis for this is the self-similarity property of the process. Traditional GARCH-type volatility models do not show this property.
- LRD models are usually fitted over larger samples and show better resistance against structural breaks. Traditional GARCH models have to be re-calibrated more frequently.

With the credit return model, future scenarios for three maturities can be simulated. With the price of the one-month maturity as a fixed point and cubic spline interpolation, the scenarios can then be applied to any maturity.

The credit return model is driven by three known risk factors:

1. The returns of a treasury bond with equal maturity.
2. A *common* credit risk factor. This describes the behavior of risky debt that is common to all corporate bonds from AAA to B in this model. After stripping off the impact of interest risk (represented by the treasury bond returns), it becomes visible that the daily movements of credit spreads of different rating grades are highly positively correlated.
3. An individual credit risk factor for each rating grade. This factor accounts for the specific risk of a given rating grade.

Individual corporate bonds, of course, do not necessarily follow the average price paths of their rating grades. For example, the spread of the bond might move from being close to the average of one rating grade to the average of another rating grade, while the rating grade of the bond itself does not change. In chapter 13 a model has been developed, describing the movement of the individual bond relative to the averages of the rating grades. This model is based on an autoregressive approach. It enables an individual bond to change its credit spread relative to the averages of the rating grades. Theoretically, its spread could be closer to the average spread of another rating grade than the one assigned to it. This may be the case when the market has a different assumption of an issuer's creditworthiness than its rating grade actually reflects.

Comparing the performance of VaR forecasting for the stable EWMA and the LRD model, both models turn out to be relatively precise in estimating the 95% VaR. Both models slightly overestimate the empirical 99% VaR. However, overall, the LRD model shows better unconditional coverage than the traditional stable EWMA model or alternative non-LRD models.

An interesting property of the FARIMA(0,d,0) model is its scaling property that comes from the self-similarity property. Traditional GARCH-type volatility models do not have this property. In order to obtain the volatility over a longer forecasting horizon, scaling with traditional GARCH type models would lead to incorrect results. Due to infinite lag polynomial governed forecasts, long horizon forecasts over several periods perform better than those performed with short-memory models.

Along with the LRD model, a new way to model the dependence between the innovations of the different rating grades / maturities has been introduced. The approach of a Gaussian copula with stable marginals is directly applicable to stable random vectors with skewed elements. It does not cause inconveniences that occur with the application of stable subordination via sub-Gaussian vectors.

It can be concluded that long-range dependence is present in the absolute returns of credit prices. Long-memory models are a powerful tool in forecasting as they deviate from the restriction that there can only be two sorts of processes, I(0) and I(1), and therefore allow more precise forecasting results. In this thesis a practical LRD model has been proposed in combination with the cointegrated VAR model to forecast credit returns.

A possible application of the model is suggested in chapter 13. It sets out how to describe the relation of the credit return model with individual corporate bonds.

Moreover, this chapter also provides a discussion of state of the art VaR-based portfolio optimization and presents a brief overview of this topic. While for a low number of positions simulation-based optimization procedures might work for VaR, larger problems with numerous positions necessitate sophisticated numerical tools and alternative risk measures. The improvement of such alternative risk measures will also improve VaR.

Finally, it can be concluded that the chosen FARIMA model represents an attractive alternative to the conventional short-memory models for volatility. It breaks the knife-edge choice restriction of I(0) and I(1) processes while at the same time remaining flexible and parsimonious for the multivariate case. Furthermore, a practical estimator to obtain the model's long-memory parameters under the stable assumption has been developed: the modified conditional sum of squares (CSS) estimator. The development of this estimator is another useful result since most estimators for the fractional differencing parameter are relatively complex and difficult to implement - especially under the stable assumption.

As they have demonstrated to outperform their competitors, stable long-memory models for the description of credit returns and the forecast of VaR will certainly play a greater role in the future.

Abbreviations

ACF	Autocorrelation function
ADF	Augmented Dickey-Fuller
APT	Arbitrage Pricing Theory
ARCH	Autoregressive Conditional Heteroskedastic (Model)
ARMA	Autoregressive Moving Average (Model)
BIS	Bank for International Settlements
BLUE	Best Linear Unbiased Estimator
CAPM	Capital Asset Pricing Model
CDF	Cumulative density function
CLT	Central Limit Theorem
CML	Capital Market Line
CSS	Conditional sum of squares
CVaR	Conditional VaR
DGP	Data-generating process
DM	Default Mode
ECM	Error Correction Model
EMH	Efficient Market Hypothesis
FARIMA	Fractional ARIMA
FIGARCH	Fractional GARCH
GARCH	Generalized Autoregressive Conditional Heteroskedastic (Model)
HMSE	Heteroscedasticity-adjusted MSE
H-ss	Self-similar with self-similarity parameter H
H-sssi	Self-similar with self-similarity parameter H and has stationary increments
i.i.d.	Independent identically distributed
IGARCH	Integrated GARCH
LL	Logarithmic Loss
LRD	Long-range dependence

MAE	Mean Absolute Error
MSE	Mean Squared Error
ML	Maximum Likelihood
MTM	Marked to Market
OECD	Organization for Economic Cooperation and Development
OLSE	Ordinary Least Squares Estimator
PACF	Partial autocorrelation function
RBC	Risk-based capital
SACF	Sample autocorrelation function
SML	Security Market Line
SPACF	Sample partial autocorrelation function
VaR	Value at Risk
VAR	Vector Autoregression
VECM	Vector Error Correction Model

Bibliography

K. Abadir and A. Taylor (1999): On the Definitions of Cointegration. *Journal of Time Series Analysis* 20: 129–137.

H. Akaike (1973): Information Theory and an Extension of the Maximum Likelihood Principle. *In: B.N. Petrov and F. Csáki (editors), 2nd International Symposium on Information Theory, Budapest, Akadémiai Kiadó* : 267–281.

H. Akaike (1974): A New Look at Statistical Model Identification. *IEEE Transactions on Automatic Control* 19: 716–723.

C. Albanese, K. Jackson and P. Wiberg (2001): Hedging With Value at Risk. *Working Paper. University of Toronto* .

J. Andel (1986): Long Memory Time Series Models. *Kybernetika* 22: 105–123.

T. Andersen and T. Bollerslev (1997): Answering the Critiques: Yes, ARCH Models Do Provide Good Volatility Forecasts. *Manuscript, Northwestern University and University of Virginia* .

T. Andersen, T. Bollerslev, F. Diebold and P. Labys (1999): Realized Volatility And Correlation. *Working Paper* .

D. Andrews (1991): Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59: 817 – 858.

P. Artzner, F. Delbaen, J. Eber and D. Heath (1998): Thinking Coherently. *Risk* 10: 68–71.

P. Artzner, F. Delbaen, J. Eber and D. Heath (1999): Coherent Measures of Risk. *Mathematical Finance* 9: 203–228.

A. Arvanitis, C. Browne, J. Gregory and R. Martin (1998): Credit Loss Distribution and Economic Capital. *Research Paper, Quantitative Credit and Risk Research, Paribas* .

L. Bachelier (1900): Theorie de la Spéculation. *Annales de Ecole Normale Superieure Series. 1964 English Translation by Coonter, P.H. (editor). The Random Character of Stock Market Prices. MIT Press, Cambridge MA* 3, 17: 28–86.

R. Baillie (1996): Long Memory Processes and Fractional Integration in Economics. *Journal of Econometrics* Vol. 73: 5–59.

R. Baillie, T. Bollerslev and H. Mikkelsen (1996): Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal Of Econometrics* 74: 3–30.

Basle-Committee (1988): International Convergence of Capital Measurement and Capital Standards. *Basle Committee on Banking Supervision. Bank for International Settlements* .

Basle-Committee (1999): Credit Risk Modeling: Current Practices and Applications. *Basle Committee on Banking Supervision. Bank for International Settlements* .

Basle-Committee (2001): The New Basle Capital Accord. *Basle Committee on Banking Supervision. Bank for International Settlements* .

M. Beck (2001): Untersuchung des Zusammenhangs zwischen Spread und Credit Ratings bei Corporate Bonds. *Seminar Work. Chair of Statistics and Econometrics. University of Karlsruhe* .

J. Beran (1994): *Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability 61*. Chapman and Hall.

J. Berkowitz (1999): Evaluating the Forecasts of Risk Models. *FEDS Discussion Paper No. 11, Federal Reserve Board of Governors* .

J. Berkowitz, I. Birgean and L. Kilian (1999): On the Finite-Sample Accuracy of Non-Parametric Resampling Algorithms for Economic Time Series. *Working Paper* .

E. Berndt, B. Hall, R. Hall and J. Hausman (1974): Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement* Vol. 3, No. 4: 653–665.

T. Bollerslev (1986): Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31: 307–327.

T. Bollerslev (1990): Modelling The Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized GARCH Approach. *Review of Economics and Statistics* 72: 498–505.

T. Bollerslev (1992): ARCH Modeling in Finance. *Journal of Econometrics* 52, (1-4).

T. Bollerslev, R. Chou and R. Engle (1992): ARCH Modeling in Finance: A Review of Theory and Empirical Evidence. *Journal of Econometrics* 52: 5 – 59.

T. Bollerslev and E. Ghysels (1996): Periodic Autoregressive Conditional Heteroskedasticity. *Journal of Business and Economic Statistics* 14.

T. Bollerslev and H. Mikkelsen (1996): Modeling and Pricing Long Memory in Stock Market Volatility. *Journal of Economics* 73: 151–184.

T. Bollerslev and J. Wooldridge (1992): Quasi-Maximum Likelihood Estimation And Inference In Dynamic Models With Time-Varying Covariances. *Econometric Reviews* 11: 143 – 172.

G. Box and G. Jenkins (1976): *Time Series Analysis: Forecasting and Control*. Prentice Hall.

Bravo-Group (2001): *Mercury 1.5 Software - Technical Document*. Bravo Risk Management Group.

Bravo-Group (2002): *Cognity Software - Technical Document*. Bravo Risk Management Group.

P. Brockwell and R. Davis (1991): *Time Series: Theory and Methods*. 2nd ed. Springer, New York.

C. Brunetti and C. Gilbert (2000): Bivariate FIGARCH And Fractional Cointegration. *Journal of Empirical Finance* Vol. 7: 509–530.

R. Burden and J. Faires (1997): *Numerical Analysis. Sixth Edition*. Brooks/Cole Publishing Company.

J. Campbell, A. Lo and A. MacKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press. Princeton, New Jersey.

W. Charemza and D. Deadman (1997): *New Directions In Econometric Practice. 2nd ed.*. Edward Elgar Publishing, Inc.

T.-L. Chong (2000): Estimating The Differencing Parameter Via The Partial Autocorrelation Function. *Journal of Econometrics* Vol. 97: 365–381.

T. Chordia, R. Roll and A. Subrahmanyan (2000a): Market Liquidity and Trading Activity. *Journal of Finance* 56, 2: 501–530.

T. Chordia, R. Roll and A. Subrahmanyan (2000b): Commonality in Liquidity. *Journal of Financial Economics* 56, 1: 3–28.

P. Christoffersen, F. Diebold and T. Schürmann (1998): Horizon Problems And Extreme Events In Financial Risk Management. *Wharton School Center for Financial Institutions, University of Pennsylvania* .

C.-F. Chung and R. Baillie (1993): Small Sample Bias In Conditional Sum Of Squares Estimators Of Fractionally Integrated ARMA Models. *Journal of Empirical Economics* Vol. 18: 791–806.

P. Cizeau, M. Potters and J.-P. Bouchaud (2000): Correlations of Extreme Stock Returns Within a Non-Gaussian One-Factor Model. *Working Paper* .

R. Cont, M. Potters and J.-P. Bouchaud (1992): Scaling in Stock Market Data: Stable Laws and Beyond. *Service de Physique de lEtat Condens. Online Publication* .

H. Demsetz (1968): The Cost of Transacting. *Quarterly Journal of Economics* 82 (February): 33–53.

D. Dickey and W. Fuller (1979): Distribution of The Estimators for Autoregressive Time Series with A Unit Root. *Journal of the American Statistical Association* 74: 427–431.

D. Dickey and W. Fuller (1981): Likelihood Ratio Statistics for Autoregressive Time Series with A Unit Root. *Econometrica* 49: 1057–1072.

Z. Ding and C. Granger (1996): Modeling Volatility Persistence of Speculative Returns. A New Approach. *Journal of Econometrics* 73: 185–215.

G. Duffee (1998): The Relation Between Treasury Yields And Corporate Bond Yield Spreads. *Journal Of Finance* 53: 2225–2241.

B. Efron (1979): Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics* 7: 1–26.

J. Elder and P. Kennedy (2001): Testing for Unit Roots: What Should Students Be Taught? *Journal of Economic Education* 31, 2: 137–146.

P. Embrechts, F. Lindkog and A. McNeil (2001): Modelling Dependencies with Copulas and Applications to Risk Management. *Working Paper, Department of Mathematics, ETH Zurich* .

P. Embrechts, A. McNeil and D. Straumann (1999): Correlation and Dependence in Risk Management: Properties and Pitfalls. *Working Paper, Department of Mathematics, ETH Zurich* .

R. Engle (1982): Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica* 50: 987–1008.

R. Engle and T. Bollerslev (1986): Modelling the Persistence of Conditional Variances. *Econometric Reviews* 5: 1–50.

R. Engle and C. Granger (1987): Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica* 55: 251–276.

R. Engle and K. Kroner (1995): Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11: 122–150.

E. Fama (1965a): The Behavior of Stock Market Prices. *Journal of Business* 38: 34–105.

E. Fama (1965b): Portfolio Analysis in a Stable Paretian Market. *Management Science* 11: 404 – 419.

E. Fama (1970): Risk, Return, and Equilibrium. *Journal of Political Economy* 78: 30–55.

Federal-Reserve-System-Task-Force (1998): Credit Risk Models at Major US Banking Institutions: Current State of the Art and Implications for Assessment of Capital Adequacy. *Federal Reserve System Task Force on Internal Credit Models* .

W. Feller (1971): *An Introduction to Probability Theory and its Application. 2nd Edition*. Wiley, New York.

L. Fisher (1959): Determinants Of Risk Premiums On Corporate Bonds. *The Journal Of Political Economy* 67: 219–237.

R. Fox and M. Taqqu (1986): Large-Sample Properties Of Parameter Estimates For Strongly Dependent Stationary Gaussian Time Series. *The Annals of Statistics* 14, 2: 517–532.

K. French (1980): Stock Returns and the Weekend Effect. *Journal of Financial Economics* 8: 55–69.

A. Gaivoronsky and G. Pflug (2001): Value at Risk In Portfolio Optimization: Properties And Computational Approach. *Working Paper. Norwegian University of Science and Technology, and University of Vienna* : 1–27.

M. Gallagher (1998): *Fitting ARMA Models to Heavy Tailed Data*. PhD-thesis, University of California, Santa Barbara.

B. Gamrowski and S. Rachev (1994): Stable Models in Testable Asset Pricing. *In: G. Anastassiou and S.T. Rachev, editors: Approximation, Probability, and Related Fields* : 223–236.

J. Geweke and S. Porter-Hudak (1983): The Estimation and Application of Long-Memory Time Series Models. *Journal of Time Series Analysis* 4: 221–238.

M. Gibson and B. Boyer (1999): Evaluating Forecasts of Correlation Using Option Pricing. *International Finance Discussion Papers. The Federal Reserve Board* .

M. Gibson, B. Boyer and M. Loretan (1999): Pitfalls in Tests for Changes in Correlations. *IFS Discussion PaperNo. 597, Federal Reserve Board* .

L. Giraitis, P. Kokoszka and R. Leipus (2000): Stationary ARCH Models: Dependence Structure and Central Limit Theorem. *Journal of Econometric Theory* 16, 1: 3–22.

G. Götzenberger, S. Rachev and E. Schwartz (2000): Performance Measurements: The Stable Paretian Approach. *Applied Mathematics Reviews* 1: 329 – 406.

C. Granger (1986): Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economic Variables* 48: 221 – 238.

C. Granger and R. Joyeux (1980): An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* 1: 15–39.

G. Gupton, C. Finger and M. Bhatia (1997): *CreditMetrics - Technical Document*. J.P. Morgan Risk Metrics Group. New York.

P. Hall, J. Horowitz and B. Jing (1995): On Blocking Rules For the Bootstrap With Dependent Data. *Biometrika* 82: 561–574.

J. Hamilton (1994): *Time Series Analysis*. Princeton University Press.

E. Hendricks (1996): Evaluation of Value-at-Risk Models Using Historical Data. *Federal Reserve Bank of New York Economic Policy Review* 2.

J. Hosking (1981): Fractional Differencing. *Biometrika* 68.

J. Hosking (1984): Modeling Persistence in Hydrological Time Series Using Fractional Differencing. *Water Resources Research* 20: 1898–1908.

H. Huang (2000): Integrated Risk Management: "Tie the Knot". *Credit Risk Special Report. RISK March 2000* : 20–22.

G. Huberman (1982): A Simple Approach to Arbitrage Pricing Theory. *Journal of Economic Theory* 28: 183–191.

G. Huberman and D. Halka (1999): Systematic Liquidity. *Paine Webber Working Paper Series in Money, Economics, and Finance. Columbia Business School, Columbia University* .

J. Hull (2000): *Options, Futures, and Other Derivatives*. Prentice-Hall International, Inc.

H. Hurst (1951): Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers* 116: 770–808.

S. Hurst, E. Platen and S. Rachev (1997): Subordinated Market Index Models: A Comparison. *Financial Engineering and the Japanese Markets* 4: 97 – 124.

D. Isakov and C. Perignon (2000): On the Dynamic Interdependence of International Stock Markets: A Swiss Perspective. *Swiss Journal of Economics and Statistics* 136, 2: 123–146.

R. Jarrow and D. V. Deventer (1999): *Practical Usage of Credit Risk Models in Loan Portfolio and Counterparty Exposure Management*. RISK Publications.

M. Jensen (1968): The Performance of Mutual Funds in the Period 1945-1964. *Journal of Finance* 23.

S. Johansen (1988): Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12: 231–254.

S. Johansen (1991): Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59: 1551 – 1580.

S. Johansen (1995): *Likelihood-Based Inference In Cointegrated Vector Autoregressive Models*. Oxford University Press, New York.

C. Jones, O. Lamont and R. Lumsdaine (1998): Public Information and the Persistence of Bond Market Volatility. *Journal of Financial Economics* 47, 3: 315–337.

S. Katz (1974): The Price Adjustment Process Of Bonds To Rating Classifications. *The Journal Of Finance* 29: 551–559.

S. Kealhofer, S. Kwok and W. Weng (1998): Uses And Abuses Of Bond Default Rates. *KMV Corporation* .

I. Khindanova, S. Rachev and E. Schwartz (1999): Stable Modeling of Value at Risk. *To appear in "Stable Models in Finance". Editors Rachev S.T. and S. Mittnik. Pergamon Press* .

P. Kokoszka and M. Taqqu (1996): Parameter Estimation For Infinite Variance Fractional ARIMA. *The Annals of Statistics* 24, 5: 1880 – 1913.

C. Lamoreux and W. Lastrapes (1990): Persistence in Variance, Structural Change, and the GARCH Model. *Journal of Business and Economic Statistics* 8: 225–235.

J. Lintner (1965): Security Prices, Risk, and Maximal Gains from Diversification. *Journal of Finance* 20.

A. Lo (1991): Long-Term Memory in Stock Market Prices. *Econometrica* 59, 5: 1279 – 1313.

F. Longin (1999b): Stress Testing: A Method Based on Extreme Value Theory. *Research Paper, The Third Annual BSI Gamma Foundation Conference* .

F. Longin and B. Solnik (1995): Is Correlation in International Equity Returns Constant: 1960-1990? *Journal of International Money and Finance* 14, 1: 3–26.

F. Longin and B. Solnik (1999): Correlation Structure of International Equity Markets During Extremely Volatile Periods. *Working Paper* .

F. Longin and B. Solnik (2001): Extreme Correlation Of International Equity Markets. *Journal of Finance* 26, No. 2.

F. Longstaff and E. Schwartz (1995): Valuing Credit Derivatives. *Journal of Fixed Income* Vol. 5: 7–12.

J. Lopez (1999): Methods for Evaluating Value-at-Risk Estimates. *FRBSF Economic Review No 2* 2.

J. Lopez and C. Walter (2000): Is Implied Correlation Worth Calculating: Evidence from FForeign Exchange Option Prices. *Journal of Derivatives* 7, 3.

J. Lopez and C. A. Walter (2001): Evaluating Covariance Matrix Forecasts in a Value-at-Risk Framework. *Journal of Risk* 3, 3: 69–98.

H. Lütkepohl (1994): Kointegration und gemeinsame Trends. *Working Paper* .

H. Lütkepohl (1998): A Review of Systems Cointegration Tests. *Working Paper, Sonderforschungsbereich 373, Humboldt University Berlin* .

H. Lütkepohl and J. Breitung (1996): Impulse Response Analysis Of Vector Autoregressive Process. *Working Paper, Sonderforschungsbereich 373, Humboldt University Berlin* .

G. Maddala and I. Kim (1998): *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press.

E. Maharaj (1999): A Test For The Difference Parameter Of The ARFIMA Model Using The Moving Blocks Bootstrap. *Working Paper 11/99. Department of Econometrics and Business Statistics, Monash University* .

B. Mandelbrot (1963): The Variation of Certain Speculative Prices. *Journal of Business* 36: 394–419.

B. Mandelbrot (1997a): *Fractals and Scaling in Finance, Discontinuity, Concentration, Risk*. Springer. New York, Berlin, Heidelberg.

B. Mandelbrot (1997b): Fractals, Hasard, et Finance. *Flammarion, Paris* .

B. Mandelbrot (1999): Survey of Multifractality in Finance. *Cowles Foundation Discussion Paper No. 1238, Yale University, New Haven* .

B. Mandelbrot and J. V. Ness (1968): Fractional Brownian Motions, Fractional Brownian Noises and Applications. *SIAM Review* 10: 422–437.

B. Mandelbrot and E. Wallis (1968): Noah, Joseph, and Operational Hydrology. *Water Resources Research* 4: 909–918.

P. Mansfield, S. Rachev and G. Samorodnitsky (1999): Long Strange Segments of a Stochastic Process and Long Range Dependence. *To appear in: The Annals of Applied Probability* .

C. Marinelli, S. Rachev and R. Roll (2001): Subordinated Exchange Rate Models: Evidence for Heavy-Tailed Distributions and Long-Range Dependence. *Mathematical and Computer Modelling* 34: 9–11.

C. Marinelli, S. Rachev, R. Roll and H. Göppl (1999): Subordinated Stock Price Models: Heavy Tails and Long-Range Dependence in the High-Frequency Deutsche Bank Price Record. *in G. Bol (editor), Data Mining and Computational Finance, Springer* .

H. Markowitz (1952): Portfolio Selection. *Journal of Finance* 8: 77–91.

B. Martin and S. Rachev (2001): A Stable Cointegrated VAR Model For Credit Returns With Time-Varying Volatility. *To appear in Proceedings of the IFAC Symposium on Modeling and Control of Economic Systems (SME) 2001* .

B. Martin, S. Rachev and E. Schwartz (2002): Stable Non-Gaussian Models in Credit Risk Management. *To appear in the Handbook of Heavy Tailed Distributions in Finance, North Holland Handbooks of Finance (Series Editor W. T. Ziemba)* .

H. Mausser and D. Rosen (1998): Beyond VaR: From Measuring Risk To Managing Risk. *Algo Research Quarterly* 1, 2: 5–20.

H. Mausser and D. Rosen (1999): Applying Scenario Optimization to Portfolio Credit Risk. *Algo Research Quarterly* 2, 2: 19–34.

H. McCulloch (1971): Measuring the Term Structure of Interest Rates. *Journal of Business, Issue 1* 44: 19–31.

H. McCulloch (1975): The Tax-Adjusted Yield Curve. *Journal of Finance* 30, 3: 811–830.

J. McKinnon (1991): Critical Values for Co-Integration. *In R.F. Engle and C.W.J. Granger (editors): Long-Run Economic Relationships. Oxford University Press* : 267 – 276.

A. McLeod and K. Hipel (1978): Preservation of the Rescaled Adjusted Range, 1: A Reassessment of the Hurst Phenomenon. *Water Resources Research* 14: 491 – 508.

R. Merton (1974): On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance* 29, 2.

T. Mikosch and C. Starica (2000a): Change of Structure in Financial Time Series, Long Range Dependence and The GARCH Model. *Working Paper. The Wharton School of Business and Chalmers University of Technology* .

T. Mikosch and C. Starica (2000b): Long Range Dependence Effects And ARCH Modelling. *Working Paper. University of Groningen and Chalmers University* .

S. Mittnik, S. Rachev and M. Paolella (1997): Stable Paretian Models in Finance: Some Theoretical and Empirical Aspects. *Appeared in: A Practical Guide to Heavy Tails. Statistical Techniques for Analyzing Heavy Tailed Distributions, editors: R. Adler et al.* .

M. Ong (1999): *Internal Credit Risk Models. Capital Allocation and Performance Measurement*. Risk Books by Risk Publications, London.

M. Ooms and J. Doornik (1999): Inference And Forecasting For Fractional Autoregressive Moving Average Models, With An Application To US And UK Inflation. *Econometric Institute Report 9947/A* .

S. Ortobelli, I. Huber and M. Höchstötter (2001): A Comparison Among Gaussian and Non-Gaussian Portfolio Choice Models. *To appear in Proceedings of the IFAC Symposium on Modeling and Control of Economic Systems (SME) 2001* .

M. Osterwald-Lenum (1992): A Note with Fractiles of the Asymptotic Distribution of the Maximum-Likelihood Cointegration Rank Test Statistics: Four Cases. *Oxford Bulletin of Economics and Statistics* 54: 461–472.

J. Park and P. Philipps (1988): Statistical Inference in Regressions with Integrated Processes: Part I. *Econometric Theory* 4: 468 – 497.

S. Park (1997): Cointegration and Market Efficiency: An Application to the Canadian Treasury Bill Market. *Carleton Economic Papers, Carleton University, Department of Economics* .

E. Peters (1994): *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. John Wiley & Sons.

P. Philipps and M. Loretan (1991): Estimating Long-Run Economic Equilibria. *Review of Economic Studies* 58: 407 – 436.

R. Pindyck and D. Rubinfeld (1991): *Econometric Models And Economic Forecasts. 3rd edition*. McGraw-Hill.

S. Rachev and S. Mittnik (2000): *Stable Paretian Models in Finance*. John Wiley & Sons, Series in Financial Economics.

S. Rachev, E. Schwartz and I. Khindanova (2001): Stable Modeling Credit Risk. *Technical Report. University of California at Santa Barbara* .

P. Robinson (1994): Semiparametric Analysis of Long Memory Time Series. *The Annals of Statistics* 22: 515–539.

P. Robinson and D. Marinucci (1998): Semiparametric Frequency Domain Analysis of Fractional Cointegration. *Discussion Paper No. EM/98/348, LSE, London* .

R. Roll (1977): A Critique of the Asset Pricing Theory's Tests: Part I on Past and Potential Testability of the Theory. *Journal of Financial Economics* 4: 129–176.

R. Roll (1978): Ambiguity When Performance is Measured by the Security Market Line. *Journal of Finance* 33: 1051–1069.

R. Roll (1989): Price Volatility, International Market Links, and their Implications for Regulatory Policies. *Journal of Financial Services Research* 3: 211–246.

S. Ross (1976): The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* 13: 341–360.

S. Ross (1978): Mutual Fund Separation in Financial Theory - the Separating Distributions. *Journal of Economic Theory* 17: 254 – 286.

S. Said and D. Dickey (1984): Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order. *Biometrika* 71: 599–607.

P. Saikkonen (1991): Asymptotically Effiencient Estimation of Cointegrating Regressions. *Econometric Theory* 7: 1 – 21.

G. Samorodnitsky and M. Taqqu (1994): *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, London.

A. Saunders (1999): *Credit Risk Measurement. New Approaches to Value at Risk and Other Paradigms*. Wiley Frontiers in Finance, Wiley & Sons.

M. Scheicher (2001): The Comovements of Stock Markets in Hungary, Poland, and the Czech Republic. *International Journal of Finance & Economics* 6, 1: 27–39.

G. Schwarz (1978): Estimating the Dimension of a Model. *Annals of Statistics* 6: 461–464.

W. Sharpe (1964): Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance* 19: 425–442.

F. Sowell (1992): Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics* 53: 165–188.

J. Stock and M. Watson (1988): Variable Trends in Economic Time Series. *Journal of Economic Perspectives* 2: 147–174.

V. Teverovsky, M. Taqqu and W. Willinger (1995): Estimators for Long-Range Dependence: An Empirical Study. *Fractals* 3, 4: 785 – 788.

V. Teverovsky, M. Taqqu and W. Willinger (1998): A Critical Look at Lo's Modified R/S Statistic. *To appear in Journal of Statistical Planning and Inference* .

G. Teyssiere (2000): Modelling Exchange Rate Volatility With Multivariate Long-Memory ARCH Process. *The European Commission, Joint Research Centre & GREQAM, Marseille. GREQAM DT 97B03* .

Y. Tokat, S. Rachev and E. Schwartz (2001): The Stable Non-Gaussian Asset Allocation: A Comparison With the Classical Gaussian Approach. *To appear in Journal of Economic Dynamics and Control* .

J. Wang (2000): Mean-Variance-VaR Based Portfolio Optimization. *Working Paper. Valdosta State University* .

W. Willinger, M. Taqqu and A. Erramilli (1996): A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High Speed Networks. *Appeared in: Stochastic Networks: Theory and Applications. F.P. Kelly, S. Zachary, and I. Ziedins, editors* .

List of Figures

List of Tables